A FAULT DETECTION AND DIAGNOSIS METHODOLOGY FOR RAILWAY TURNOUT SYSTEMS BASED ON PARAMETER ESTIMATION

by

TANAPON KUMPAO

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Electronic, Electrical, and Systems Engineering

College of Engineering and Physical Sciences

School of Engineering

University of Birmingham

May 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

This page is left intentionally blank in anticipation of double-sided printing.

ABSTRACT

Railway turnout systems are one of the most important signalling elements in railway operations, such that their failure during service hours can adversely impact service reliability and safety. Thus, effective maintenance must be applied to ensure that their failure is prevented. For such purposes, data-driven fault detection and diagnosis (FDD) methodologies have been commonly proposed as viable solutions. However, regardless of their exceptional accuracy, they usually employ advanced data analysis algorithms which are difficult for maintenance staff to comprehend. Moreover, their overfitting nature requires them to undergo a sophisticated re-training process before they can be effectively transferable to diagnose other railway turnout systems.

Due to these difficulties in their practical applications, this thesis proposes an FDD methodology developed from a parameter estimation technique instead, where underlying system knowledge is used to simplify diagnostic reasoning. Once the proposed FDD methodology is applied to a real turnout, it can estimate physically interpretable turnout parameters. These parameters exhibit changes in patterns when a particular fault mode is present in the turnout system. Due to the interpretability of estimation results and specific changes in the parameters, an interpretable and comprehensible classifier can consequently be developed.

The classifier achieves more than 80% overall accuracy when it is applied to the training turnout system. For every diagnostic decision, the skilled personnel can also re-assess it based on the interpretable parameter estimates used in the classification. However, when the proposed methodology is immediately applied to diagnose a different turnout system, it only achieves approximately 60% overall accuracy. Due to the interpretable and traceable diagnostic reasoning used in the classifier, a simplified cross-training and tuning procedure can be implemented. This improves the accuracy to nearly 90% in diagnosing the test turnout system.

In summary, this thesis proposes and develops a parameter estimation-based FDD methodology for railway turnout system applications, which realises interpretability and traceability as well. Achieving transferability is also possible, but the FDD methodology must undergo some degree of cross-training first for its application to be effective on different turnout systems.

ACKNOWLEDGEMENT

This thesis could never have been completed without the guidance of my supervisors, Dr. Edward Stewart and Prof. Roger Dixon. I am profoundly grateful for all the support and positive relationships we have shared over the years, and I hope they will endure for many more. I extend my heartfelt thanks to Adnan Zentani, Dr. Saikat Dutta, and Dr. Paul Weston for their assistance with the technical issues I encountered during my experiments; to John O'Neill for providing practical information on the HW55S turnout system; to BK3 Training Limited for allowing me to conduct experiments on the HW2000 turnout system, which helped validate the methodology's transferability; and to Dr. Ramakrishnan Ambur, Dr. Osama Olaby, Dr. Holly Foss, all BCRRE staff, and all my PhD colleagues, whom I had the pleasure of meeting during my studies.

I would also like to express my gratitude to Dr. Hongsin Kim for granting me the opportunity to serve as a postgraduate teaching assistant in her courses. This experience has prepared me for the lecturer position I will pursue in Thailand and simultaneously provided financial support. I am thankful as well for the support of Prof. Felix Schmid, Dr. Ning Zhao, Stephen Kent, and Prof. Stuart Hillmansen, to whom I extend my blessings. Lastly, I am grateful to all the staff at Chaophraya Birmingham restaurant for their Thai hospitality, which made me feel as though I had never truly left Thailand, especially P'Pop and P'A, who provided me with accommodation during my time of greatest need.

This opportunity would not have been possible without the Royal Thai Government Scholarship, which provided financial support throughout my studies in the UK. My decision to study was greatly supported by Associate Prof. Mongkol Konghirun, who had genuine faith in me as a valuable addition to the Electrical Engineering Department at King Mongkut's University of Technology Thonburi.

Finally, no support compares to that of family. I am deeply indebted to my wife, Pim, who has stood by me through thick and thin, through COVID-19, through times of joy, and through times of sorrow. I eagerly anticipate the next chapter that this doctoral degree will bring us to. Last but certainly not least, I am grateful for our two amazing cats, Pumpkiin, the ginger, and Bibi, the black, to whom I am committed to providing the best life imaginable.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

1.1 Problem statement 2

1.2 Aim and objectives 5

1.3 Scope of the thesis 6

1.4 Thesis structure 6

CHAPTER 2 BACKGROUND OF RAILWAY TURNOUTS AND LITERATURE REVIEW 10

2.1 Background of railway turnout systems 10

2.1.1 Functional procedures of railway turnout systems 10

2.1.2 Common types of railway point machines 12

2.1.3 Operational behaviour of electro-mechanical turnout systems 14

2.1.4 Common fault modes of electro-mechanical point machines 20

2.1.5 Conclusions 33

2.2 Literature review of railway turnout FDD developments 33

2.2.1 Data-driven methodologies 35

2.2.2 Model-based methodologies 52

2.2.3 Summary of limitations of existing FDD methodologies 60

2.2.4 Proposed feature extraction technique: Parameter estimation 63

2.2.5 Conclusions 67

CHAPTER 3 APPLICATION OF PARAMETER ESTIMATION FOR RAILWAY TURNOUT FDD 69

3.1 Parameter estimation for FDD purposes 69

TABLE OF CONTENTS (Cont.)

3.2 Overview of parameter estimation algorithms ..……..……71

3.2.1 Indirect parameter estimation principle ……..……...71

3.2.2 Direct parameter estimation principle …….……75

3.2.3 Comparison between indirect and direct parameter estimation ……..……77

3.3 The SRIVC parameter estimation algorithm ..……..…...……78

3.3.1 The solution to asymptotic bias: instrumental variables (IVs) ... 79

3.3.2 Reduction of numerical differentiation errors by using pre-filters …..…..…… 80

3.3.3 Procedural steps of the SRIVC algorithm ..……...…… 82

3.4 Simulation model of DC electro-mechanical turnouts ...…… 85

3.5 Application of the SRIVC algorithm to DC railway turnouts ..…… 90

3.6 Verification of the SRIVC algorithm performance ...…..……97

3.7 Conclusions ……..…….....104

CHAPTER 4 VALIDATION OF THE PROPOSED TRANSFER FUNCTION MODELS TO REPRESENT SWITCH RAIL DYNAMICS ……...106

4.1 Introduction to switch rail bending dynamics ……..…… 106

4.2 Validation of the turnout's transfer function models ..…… 112

4.3 Validation of the turnout's transfer function models ..…… 115

4.3.1 Motor winding fault simulation ……..……...……...116

4.3.2 Wear and tear in the point machine fault simulation ...……117

4.3.3 Physically-related motor fault simulation ……..……...118

4.3.4 Improper sliding surface fault simulation ……..……119

4.4 Conclusions ……..…….....120

TABLE OF CONTENTS (Cont.)

CHAPTER 5 PARAMETER CHANGE DETECTION CONDUCTED ON AN HW55S TURNOUT SYSTEM WITH SIMULATED FAULTS 122

5.1 HW55S turnout system and experimental setup 123

5.2 Fault-free conditions of HW55S turnout 128

5.2.1 Modifications of SRIVC parameter estimation..... 131

5.2.2 Baseline of all fault-free HW55S turnout parameters..... 136

5.3 Fault simulations imposed onto the HW55S turnout 140

5.3.1 Overdrive fault simulations 141

5.3.2 Obstruction fault simulations 147

5.3.3 Underdrive fault simulations 153

5.3.4 Abnormal input voltage fault simulations 159

5.3.5 Improper sliding surface fault simulations 162

5.4 Parameter change “knowledge-base” table 166

5.5 Conclusions 168

CHAPTER 6 DEVELOPMENT OF FAULT DETECTION AND DIAGNOSIS ALGORITHM 170

6.1 Rule-based FDD classification algorithm 170

6.2 Fuzzy logic rule-based classification..... 171

6.2.1 Introduction of fuzzy logic 171

6.2.2 Fuzzification of turnout parameter estimates 173

6.2.3 Formulation of fault detection rule in the fuzzy inference system 176

6.2.4 Fault diagnosis rules and interpretation of FDD results 182

6.3 Results from the HW55S turnout system 189

6.3.1 Results from the training dataset 189

TABLE OF CONTENTS (Cont.)

6.3.2 Results from the testing dataset …….....……....193

6.4 Transferability validation: HW2000 turnout system ……196

6.4.1 Experimental setup of the HW2000 turnout system ..……....197

6.4.2 Initial FDD results before modification ….....……....198

6.4.3 Modification of the fuzzy logic-based FDD algorithm ..……..... 203

6.4.4 Final FDD results after modification …….....……..... 207

6.5 Conclusions ……..…….....209

CHAPTER 7 CONCLUSIONS AND FUTURE WORK ……... 212

7.1 Conclusions of the FDD methodology development …..…….. 212

7.2 Recommendations and future works ……. 215

REFERENCES AND BIBLIOGRAPHY ……..... 217

APPENDIX 1 TERMINOLOGIES PREVIOUSLY USED TO DESCRIBE COMMON FAULT MODES ……..... 235

APPENDIX 2 PARAMETER ESTIMATION RESULTS OF THE HW55S TURNOUT SYSTEM ……... 237

APPENDIX 3 FUZZY FAULT DIAGNOSIS RULES ……..... 254

APPENDIX 4 DIAGNOSTIC RESULTS FROM THE HW55S TURNOUT SYSTEM ……..... 258

APPENDIX 5 DIAGNOSTIC RESULTS FROM THE HW2000 TURNOUT SYSTEM ……..... 266

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

Figure 1-1 Records of UK rail passenger journeys from 1950 to the emergence of COVID-19 in 2020–2021 and the most recent recovering trend 1

Figure 1-2 Example of the layout of a railway electro-mechanical turnout system and its associated permanent way components 2

Figure 1-3 Thesis structure and chapters with their associated objectives 9

CHAPTER 2 BACKGROUND AND LITERATURE REVIEW

Figure 2-1 High-level diagram of the turnout's dynamic "direct path of trains according to signalling command" function 11

Figure 2-2 Array of 16 sensors used to monitor an M63 turnout system 15

Figure 2-3 Times each measurement was analysed in previous railway turnout FDD developments 15

Figure 2-4 Examples of segmented current waveforms (a) S700K; it can be segmented into either three or five phases, (b) ZD6, and (c) HW55S 16

Figure 2-5 Segmentation of motor speed and power measurements from Alstom DK FBANE MET point machines into seven phases and fault simulation results 19

Figure 2-6 Illustrations of faults and failures temporal characteristics 21

Figure 2-7 Number of acknowledgements of each common fault mode of electro-mechanical turnout systems 22

Figure 2-8 General structure of a railway turnout locking mechanism operating in (a) fault-free states and (b) during obstruction failures 25

Figure 2-9 Examples of incipient improper sliding surface fault (a) a failure condition, (b) simulated by increasing the number of contaminated slide chairs, and (c) using an extra hydraulic device 27

LIST OF FIGURES (Cont.)

Figure 2-10 The driving and the “adjuster” mechanism that provides a “lost-motion” distance for HW55S point machines 29
Figure 2-11 Overdrive fault simulations with the trajectory of the switch rails 29
Figure 2-12 Underdrive fault simulations with the trajectory of the switch rails 31
Figure 2-13 High-level transformation stages in every FDD algorithm 34
Figure 2-14 Three consecutive criteria that experts have proposed for wear-related slide chair fault detection: (a) irregularity, (b) time index of the absolute maximum force difference, and (c) symmetry of the waveform 37
Figure 2-15 Similarity analysis based on the selected subsequence 40
Figure 2-16 Different degradation trends of each fault mode 44
Figure 2-17 RSE trends from two different turnouts of the same type 45
Figure 2-18 Distributions of “mean current during the moving phase” and “throw time” features extracted from two different turnout systems of the same design (differentiated by colours) when they were both in known fault-free conditions 46
Figure 2-19 Illustration of CNN’s deep architecture for image processing 48
Figure 2-20 Generic FDD methodology for several railway STME assets 54
Figure 2-21 Comparison between the measured current (the solid line) and the estimation calculated from a first principle model (the dashed line) of a P80 system .. 56
Figure 2-22 Flowchart illustrating the entire development process of the proposed parameter estimation-based FDD approach for electro-mechanical turnout systems .. 68

CHAPTER 3 APPLICATION OF PARAMETER ESTIMATION FOR RAILWAY TURNOUT FDD

Figure 3-1 Process of obtaining CT variables for direct parameter estimation 76
Figure 3-2 High-level flowchart of SRIVC algorithm procedures 85

LIST OF FIGURES (Cont.)

Figure 3-3 Schematic diagram of a permanent magnet DC simulation model that represents the linear actuation mechanism of a DC railway turnout 86

Figure 3-4 Permanent magnet DC turnout simulation model 89

Figure 3-5 Four variables collected from the DC turnout simulation in Figure 3-4, which are also compared to the measurements collected from a fault-free HW55S throw 99

Figure 3-6 Components of a boxplot that will be used to represent a distribution of estimates 100

Figure 3-7 Parameter estimation results from the LS algorithm 101

Figure 3-8 Parameter estimation results from the SRIV algorithm 101

Figure 3-9 Parameter estimation results from the SVF algorithm..... 102

Figure 3-10 Parameter estimation results from the SRIVC algorithm..... 102

CHAPTER 4 VALIDATION OF THE PROPOSED TRANSFER FUNCTION MODELS TO REPRESENT SWITCH RAIL DYNAMICS

Figure 4-1 The switch rail assembly that the point machine can only actuate at the drive stretcher bar and the switch rail's differential equation to be validated 106

Figure 4-2 The changing geometry of the switch rails that start from cross-sections identical to those of the stock rails at the heel end and become narrower towards the front toe..... 107

Figure 4-3 The co-simulation environment between SIMPACK and Simulink® using the SIMAT interface to allow data exchange between the two models 108

Figure 4-4 A 3D model of the stock rail and the switch rail demonstrating the changing of size and shape of cross sections from Point Z to the front toe 109

Figure 4-5 Schematic of the connections between two switch rails that are connected by a drive stretcher bar and being confined between two stock rails 111

LIST OF FIGURES (Cont.)

Figure 4-6 Li's FEA model that is integrated into the DC turnout simulation model in order to represent the switch rail's bending dynamics 113
Figure 4-7 Comparison between the simulation data that integrated Li's FEA model and the SRIVC estimation results, with the coefficient of determination 114
Figure 4-8 SRIVC estimation results of the measurements collected from the simulated motor winding faults by changing the armature resistance and inductance 116
Figure 4-9 SRIVC estimation results of the measurements collected from the simulated motor's wear and tear faults by changing the motor's inertia and damping..... 117
Figure 4-10 SRIVC estimation results of the measurements collected from the simulated physically related motor faults by changing the voltage constant 118
Figure 4-11 SRIVC estimation results of the measurements collected from the simulated improper sliding surface faults by changing the coefficient of friction 119

CHAPTER 5 PARAMETER CHANGE DETECTION CONDUCTED ON AN HW55S TURNOUT SYSTEM WITH SIMULATED FAULTS

Figure 5-1 Examples of HW turnout systems currently used in UK railways..... 122
Figure 5-2 Available HW55S turnout system and its connections to the switch rail assembly..... 123
Figure 5-3 Structure and components of the HW55S turnout system 124
Figure 5-4 Actuation mechanism of the HW55S turnouts operating in N-to-R 125
Figure 5-5 All sensor installations in the HW55S turnout system 127
Figure 5-6 Measurements obtained from fault-free throws of the HW55S turnout 128
Figure 5-7 Comparison between the fault-free measurements from the selected throw and the responses of the validated transfer functions with the SRIVC estimates 129
Figure 5-8 Procedural steps of the proposed segmentation algorithm..... 132

LIST OF FIGURES (Cont.)

Figure 5-9 Comparison between the measurements of the same selected throw, as shown in Figure 5-7, and the results from the estimation 137

Figure 5-10 Distributions of parameter estimates from all fault-free HW55S throws using the interquartile range as the baseline for each parameter..... 139

Figure 5-11 Simulation process of overdrive faults conducted on the HW55S..... 141

Figure 5-12 Details of HW55S unlocking mechanism by de-coupling the lock dog from the lock notch and the restoration force that makes unlocking more difficult 142

Figure 5-13 Comparison of motor current and motor speed measurements between fault-free and simulated overdrive N-to-R throws 143

Figure 5-14 Estimation results of overdrive N-to-R throws simulated in all three severity levels, with the zoom-in distributions of *TotDist* and $\omega_m(t_0)$..... 144

Figure 5-15 Comparison of motor current and motor speed measurements between fault-free and simulated unlock difficulty N-to-R throws..... 145

Figure 5-16 Estimation results of unlock difficulty N-to-R throws simulated in all three severity levels, with the zoom-in distributions of *TotDist* 146

Figure 5-17 Three different thicknesses of metal shim placed in two positions on the HW55S turnout system to simulate obstruction faults 148

Figure 5-18 Comparison of motor current and motor speed measurements between fault-free and simulated obstruction N-to-R throws 149

Figure 5-19 Estimation results of obstruction N-to-R throws simulated in three severity levels, with the zoom-in distributions of *TotDist* and $\omega_m(t_0)$..... 150

Figure 5-20 Comparison of motor current and motor speed measurements between fault-free and simulated OPP-obstruction N-to-R throws..... 151

Figure 5-21 Estimation results of OPP-obstruction N-to-R throws simulated in three severity levels, with the zoom-in distributions of *TotDist* 152

LIST OF FIGURES (Cont.)

Figure 5-22 Details of HW55S locking mechanism using the coupling of the lock dog and the lock notch and their misalignment caused by underdriving failures 154
Figure 5-23 Comparison of motor current and motor speed measurements between fault-free and simulated underdrive N-to-R throws 155
Figure 5-24 Estimation results of underdrive N-to-R throws simulated in three severity levels, with the zoom-in distributions of *TotDist* 156
Figure 5-25 Comparison of motor current and motor speed measurements between fault-free and simulated OPP-underdrive N-to-R throws 157
Figure 5-26 Estimation results of OPP-underdrive N-to-R throws simulated in three severity levels, with the zoom-in distributions of *TotDist* 158
Figure 5-27 Drive circuit from the lineside LOC to the working turnout system displaying connections in between and where the voltages can be measured 159
Figure 5-28 Comparison of motor current and motor speed measurements between fault-free and simulated abnormal input voltage N-to-R throws 160
Figure 5-29 Estimation results of abnormal input voltage N-to-R throws simulated in three severity levels, with the zoom-in distributions of k_M and $\omega_m(t_0)$ estimates 161
Figure 5-30 Samples of the fault-free slide chair condition 163
Figure 5-31 Samples of the dry slide chair condition 163
Figure 5-32 Samples of the greased slide chair condition 164
Figure 5-33 Comparison of motor current and motor speed measurements between fault-free and simulated improper sliding surface N-to-R throws 164
Figure 5-34 Estimation results of improper sliding surface N-to-R throws simulated in three slide chairs' surface conditions 165
Figure 5-35 Inspection of the interface between the switch rails and the slide chairs 166

LIST OF FIGURES (Cont.)

CHAPTER 6 DEVELOPMENT OF FDD ALGORITHM

Figure 6-1 High-level workflow diagram of the proposed fuzzy logic-based FDD algorithm 173

Figure 6-2 Distribution of unaffected B_{TO} parameter estimates obtained from N-to-R throws presented in histogram, approximated distribution curve, and boxplot 174

Figure 6-3 Distribution of other unaffected parameter estimates 175

Figure 6-4 Example of the membership functions formulation for the B_{TO} 176

Figure 6-5 Distributions of the three *TotDist* membership values from all four fault conditions that reduce the switch rail actuation stroke significantly 177

Figure 6-6 Unique effects from the N-to-R 9-face overdrive fault scenarios 178

Figure 6-7 Unique effects from the N-to-R 9-face unlock difficulty fault scenarios 179

Figure 6-8 Histograms of the “fault-free” membership values of all N-to-R fault-free throws and all N-to-R faulty throws calculated from the fault detection rule 181

Figure 6-9 Example of FDD result display panel for a fault-free N-to-R throw 186

Figure 6-10 Example of FDD result display panel for a 3-face overdrive N-to-R 187

Figure 6-11 Example of FDD result display panel for a 9-face overdrive N-to-R 187

Figure 6-12 Example of FDD result display panel for a OPP-obstruction N-to-R 188

Figure 6-13 Example of FDD result display panel for a 9-face underdrive N-to-R 188

Figure 6-14 Confusion matrix showing the FDD results of the N-to-R HW55S turnout training dataset produced by the proposed methodology 0.5 190

Figure 6-15 Examples of occurrence likelihood trends extracted from the training dataset 192

Figure 6-16 Confusion matrix showing the FDD results of the N-to-R HW55S turnout testing dataset produced by the proposed methodology 0.5 194

Figure 6-17 Fault occurrence likelihood trends of the entire N-to-R HW55S testing dataset 195

LIST OF FIGURES (Cont.)

Figure 6-18 Overview of the HW2000 turnout system at BK3 Training Limited 197

Figure 6-19 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R fault-free throw 198

Figure 6-20 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R 6-face unlock difficulty throw 199

Figure 6-21 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R 4 mm obstruction throw 199

Figure 6-22 FDD display panel of fault-free HW2000 N-to-R throw 200

Figure 6-23 FDD display panel of 6-face unlock difficulty HW2000 N-to-R throw 200

Figure 6-24 FDD display panel of 4 mm obstruction HW2000 N-to-R throw 201

Figure 6-25 Confusion matrix showing the FDD results of the N-to-R HW2000 turnout testing dataset produced by the proposed methodology 0.5 202

Figure 6-26 Fault occurrence likelihood trends of the entire N-to-R HW2000 testing dataset 202

Figure 6-27 The system overview of the proposed FDD methodology, including the cross-training process required for the methodology to be effective for another turnout system 204

Figure 6-28 Differences between the effects of the 6-face unlock difficulty N-to-R operations on the parameter estimates of HW55S and HW2000 turnout systems 205

Figure 6-29 Differences between the effects of the 4 mm obstruction N-to-R operations on the parameter estimates of HW55S and HW2000 turnout systems 206

Figure 6-30 Confusion matrix showing the FDD results of the N-to-R HW2000 turnout testing dataset once the methodology has been modified to suit the HW2000 208

Figure 6-31 Fault occurrence likelihood trends of the entire N-to-R HW2000 testing dataset once the methodology has been modified to suit the HW2000 turnout 208

LIST OF TABLES

CHAPTER 2 BACKGROUND AND LITERATURE REVIEW

Table 2-1 Some key terms used in the 35 selected publications that related to the common fault modes being categorised as in Figure 2-7 23

Table 2-2 Summary of comparison analysis on the previously developed data-driven FDD methodologies for railway turnout systems 51

Table 2-3 Summary of comparison analysis on the previously developed model-based FDD methodologies for railway turnout systems 59

Table 2-4 Evaluations of each FDD category of the three proposed requirements 62

CHAPTER 3 APPLICATION OF PARAMETER ESTIMATION FOR RAILWAY TURNOUT FDD

Table 3-1 All parameter values used in the simulation model of a permanent magnet DC electro-mechanical turnout system 97

CHAPTER 4 VALIDATION OF THE PROPOSED TRANSFER FUNCTION MODELS TO REPRESENT SWITCH RAIL DYNAMICS

Table 4-1 SRIVC estimated parameters of Eqs. (3-29) and (3-42) when Li's FEA model is used to represent the switch rail's dynamics instead of a mass-spring-damper assembly 114

CHAPTER 5 PARAMETER CHANGE DETECTION CONDUCTED ON AN HW55S TURNOUT SYSTEM WITH SIMULATED FAULTS

Table 5-1 All sensors used in the data acquisition system and their details 127

Table 5-2 SRIVC estimation results of the selected throw and the numbers in parentheses as their corresponding estimated standard deviation 130

LIST OF TABLES (Cont.)

Table 5-3 SRIVC parameter estimates of the same selected throw, as plotted in Figure 5-7, with their corresponding estimated standard deviation in parentheses 137

Table 5-4 Number of throws simulated during the training data collection campaign 140

Table 5-5 Parameter change knowledge-base table summarising the correlation between the considered fault modes and the trend of changes of every turnout parameter 167

LIST OF ABBREVIATIONS

CBM	Condition-based Maintenance
CT	Continuous Time
DT	Discrete Time
FDD	Fault Detection and Diagnosis
FIS	Fuzzy Inference System
IV	Instrumental Variable
LOC	Location Case
LS	Least Square
LTI	Linear Time-invariant
MCS	Monte-Carlo Simulation
MISO	Multiple-input Single-output
ORR	Office of Rail and Road
SISO	Single-input Single-output
SNR	Signal-to-Noise Ratio
SRIVC	Simplified Refined Instrumental Variable for Continuous systems
SVF	State-variable Filter
UK	United Kingdom

This page is left intentionally blank in anticipation of double-sided printing.

CHAPTER 1 INTRODUCTION

Several countries have been attempting to promote railways as their primary mode of public transport because of their efficiency, reliability, and environmental friendliness (State Railway of Thailand, 2019; Soares et al., 2021). The efforts have become more successful over the years, as shown by the continuous growth of rail demand (Williams, 2019). In the UK, statistics presented by the Office of Rail and Road (ORR) show that a rising trend had emerged since privatisation, with an approximate rise of 3.5% per annum (ORR, 2021). The COVID-19 pandemic, which started in 2020–2021, eventually interrupted this growing demand. Nevertheless, more recent statistics have reported a steady market recovery, as shown in Figure 1-1.

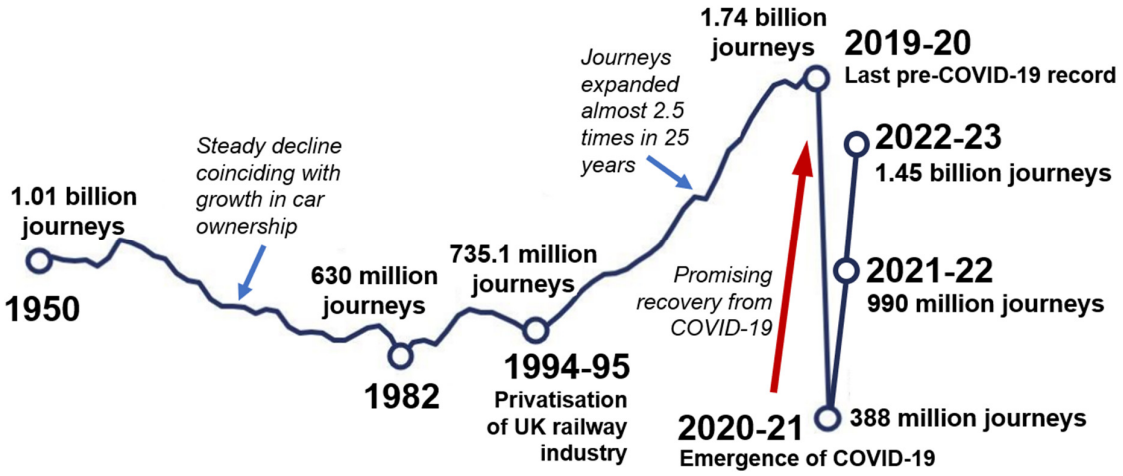


Figure 1-1 Records of UK rail passenger journeys from 1950 to the emergence of COVID-19 in 2020–2021 and the most recent recovering trend (ORR, 2021; 2023)

Similar market demand recovery has also been reported for continental Europe (SCI Verkehr GmbH, 2021), China (Gao et al., 2022), and Thailand (Department of Rail Transport, 2023). Therefore, the railway industry worldwide is still under pressure to answer these ever-increasing demands and public expectations while maintaining the highest level of safety and reliability at the same time.

Being guided transport systems, railways require movable track elements to facilitate connectivity and diverging routes for the vehicles, as they can only move in one dimension (i.e. either forwards or backwards). As a result, trains can only change their course at specific locations where these elements, called turnouts, are situated in the network (Jimenez et al., 2020). Figure 1-2 illustrates an example of their layout.

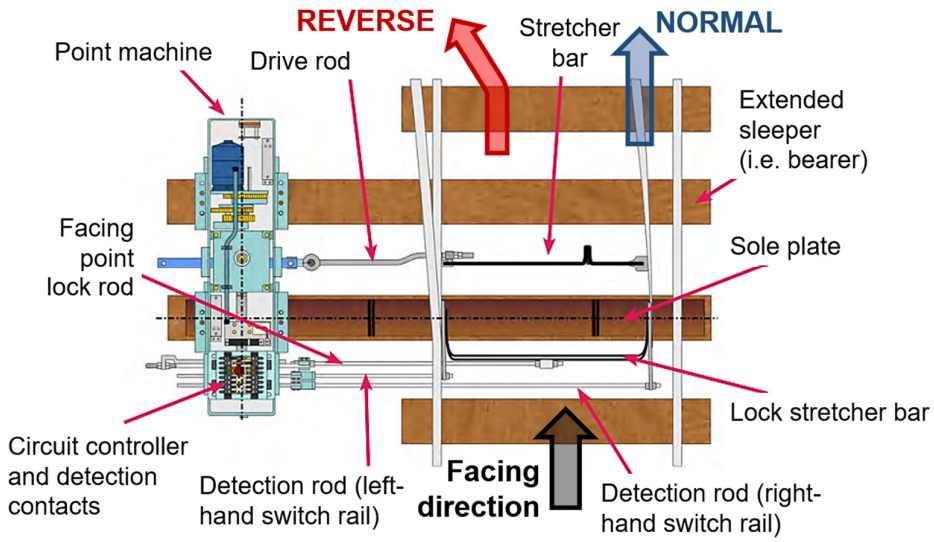


Figure 1-2 Example of the layout of a railway electro-mechanical turnout system and its associated permanent way components (How, 2022a)

Due to the functionalities of turnouts, their failures can severely affect service reliability and sometimes may cause damages and fatalities. Therefore, turnouts are considered one of the most critical railway infrastructure components and have been selected as the study subject of this thesis (Letot et al., 2015; Guzman et al., 2019).

1.1 Problem statement

Railway turnouts allow trains to be diverted along varying routes (Bell, 2005). A number of them are usually required to provide sufficient connectivity to the network. A single failure, consequently, can result in severe disruptions to the service (Jackson

and Spurlock, 2021). Note that although they are also called points or switches in other studies, the term “turnouts” will be used consistently throughout this thesis.

When turnout failures happen, signallers will lose control over them. No trains will be permitted to enter that track section to ensure the safest state, which comes from the signalling fail-safe philosophy (Zattoni, 2006). Besides service disruptions, these failure occurrences will damage the company’s reputation and may lead to fines being imposed on infrastructure operators as well (Independent Regulations’ Group – Rail, 2017). Regardless of the consequences of these failures, railway turnouts are still responsible for a large proportion of previously recorded disruptions (Zhou et al., 2001; Bemment et al., 2018; Department of Rail Transport, 2022).

Derailments can also happen if turnouts fail to lock the switch rails securely and they are moved when a train passes over. Potters Bar and Grayrigg are two of the most recognised accidents in the UK (Márquez, Schmid, and Collado, 2003; Bagwell, 2004; Rail Accident Investigation Branch, 2011). They were both found to be caused by ineffective maintenance. Due to the catastrophes resulting from these accidents, railway turnouts have been classified as safety-critical infrastructures.

As a result, railway turnouts must be appropriately inspected, monitored, and maintained to ensure the highest level of service safety and reliability (Goodwin, 1998). The combination of scheduled preventive maintenance and over-engineering designs has been applied as their solution for a long time (Roberts et al., 2002; Kun, Tao, and Cheng, 2021). However, as the previous records show, it has proven unsatisfactory.

When failures occur less frequently than the scheduled intervals, maintenance-associated resources are spent unnecessarily (Zattoni, 2006; Márquez, Weston, and Roberts, 2007). In contrast, when turnouts fail too often and happen during service

hours, they will require immediate corrective maintenance (Zhang et al., 2014; Mistry, Lane, and Allen, 2020). Costs and delays, difficulties in dispatching staff to the failed equipment's location, and safety concerns while working during service hours and adjacent to operating railways have motivated the railway industry to invest in more efficient and effective maintenance regimes (Márquez et al., 2008). Due to these reasons, the condition-based maintenance (CBM) started to emerge.

A maintenance advisory system capable of fault detection and diagnosis (FDD) can facilitate CBM, as it can *indicate whether a turnout is in its fault-free state or one of the previously acknowledged fault modes with some severity level information* (ISO, 2003). Such competences enable the maintainers to schedule maintenance for each turnout more effectively and proactively, which allows them to work more safely during engineering hours (Roberts et al., 2002; Dassanayake et al., 2009; Matei et al., 2015).

In the early development of FDD systems, thresholding methods were directly employed to compare selected system measurements with pre-determined thresholds (Zhou et al., 2002b). The user would receive an alarm if any measurement exceeded these thresholding boundaries, defined based on the experts' experience.

However, this simple approach was not successful because of its inaccuracy. When a fault is not detected and left unattended, it can become a failure during service hours. Meanwhile, false alarms will increase staff workloads and result in unnecessary expenditure of resources. Thus, users began to lose confidence in thresholding-based results and often disregarded them (Zhou et al., 2001; Dassanayake et al., 2009).

Therefore, more sophisticated ways of data analysis have been implemented to improve diagnosis accuracy. Data-driven approaches have become the main attraction in this research field since they can achieve FDD capabilities by establishing arbitrary

relationships between system measurements and their “health state” labels (Ou et al., 2019; Chen et al., 2021; Hu et al., 2022). Regardless of their exceptional accuracy, several challenges have prevented data-driven methodologies from being practical solutions for railway turnouts, including their high abstractness and overfitting nature (Silmon and Roberts, 2010; Jin et al., 2015; Panchenko et al., 2019).

There is another FDD philosophy that utilises the system’s underlying physics to develop an analytical representation, called model-based approaches (Ljung, 1999; Isermann, 2005; Ding, 2008). Such knowledge is transferrable to all system instances of the same type. Also, the classification algorithms are usually simpler to understand and more interpretable compared to those applied in the data-driven counterparts (Bai, Roberts, and Goodman, 2008). However, these methodologies have yet to be much explored for railway turnouts (Hamadache et al., 2019a; Dutta et al., 2020a). If they are proven to be viable, their interpretability and transferability may enable such model-based FDD methodologies to be practical for railway turnout applications.

1.2 Aim and objectives

This thesis aims to develop a more practical FDD methodology for railway turnout systems based on a model-based technique. The hypothesis is as follows:

“By incorporating physics knowledge of railway turnouts into model-based FDD development, a more transferable methodology that can provide sufficiently accurate, comprehensive, and interpretable diagnosis results can be established.”

To achieve this aim and evaluate the hypothesis, the following objectives must be accomplished throughout this thesis:

1. Identify a model-based technique for developing a practical FDD methodology for railway turnouts, such that the diagnosis results should be able to identify

the fault mode by an interpretable algorithm, while the proposed methodology should also be transferrable to other turnouts of the same type.

2. Assumptions and simplifications must be justified and validated by simulation studies to initially ensure the proposed model-based technique's feasibility.
3. Demonstrate the FDD capabilities of the proposed methodology by applying it to the measurements taken from an actual railway turnout system.
4. Once the performance of the proposed methodology is determined, evaluate its transferability when it is applied to another railway turnout system.

1.3 Scope of the thesis

1. The focus of this thesis will be limited to electro-mechanical turnout systems. This is because they are one of the most commonly used types in UK railways.
2. This study only considers the dynamic behaviour of railway turnouts. This is when they are instructed to actuate the switch rails by signallers. Dynamics from the substructure, passing traffic, and static properties are not evaluated.
3. Because all data will be collected from a slab-track single-turnout testbed in a laboratory, realistic railway environmental and operational variations between turnouts (e.g. temperature, humidity, type and weight of train passings, and operation frequency) will not be considered during the transferability validation.

1.4 Thesis structure

From Chapter 2 onwards, the contents of each chapter can be summarised as follows:

Chapter 2: Background of Railway Turnouts and Literature Review

This chapter will provide background knowledge of railway turnouts, including their functional procedures, dynamic behaviours, and common fault modes. Then, a literature review of previously developed FDD methodologies for railway turnouts will

be presented and assessed based on their practicality. The assessment of the current research field will lead to a proposal for *parameter estimation-based* techniques.

Chapter 3: Application of Parameter Estimation for Railway Turnout FDD

First, parameter estimation for FDD purposes will be introduced, and a model of DC turnout systems will be developed. Then, a specific algorithm will be selected based on its previous successful applications. Simulation studies will initially verify the proposed algorithm based on criteria such as the goodness of fit, statistical properties of the estimation results, noise robustness, and sampling rate sensitivity. The result will indicate if the chosen algorithm is suitable for the turnout FDD application.

Chapter 4: Validation of The Proposed Transfer Function Models to Represent Switch Rail Dynamics

The transfer function models developed in Chapter 3 will be validated in this chapter, where the switch rail's bending dynamics are considered in more detail. Once the turnout simulation model integrates realistic aspects of the bending dynamics, the performance of the selected parameter estimation algorithm and the proposed transfer functions will be validated. Afterwards, the parameter change detection capability will be verified when the simulation model is adjusted to represent faulty turnout systems.

Chapter 5: Parameter Change Detection Conducted on an HW55S Turnout System with Simulated Faults

The selected algorithm will now be validated by an actual electro-mechanical HW55S turnout. Due to the inherent non-linear dynamics of turnout operations, the parameter estimation algorithm will not be able to provide satisfactory results at first. Therefore, modifying the algorithm to address these non-linearities will be necessary.

After the modification, a fault-free range of every parameter will be defined from many known fault-free throws. Then, every fault mode considered in the Background section will be simulated on the testbed to obtain their parameter estimates. Compared to fault-free cases, there should be apparent differences in the affected parameters.

Finally, the relationships between parameter changes (detected by the modified algorithm) and their corresponding simulated fault mode will be summarised. This information will be the foundation for the FDD algorithm development in Chapter 6.

Chapter 6: Development of Fault Detection and Diagnosis Algorithm

Now, the parameter estimates will be considered as “FDD features”. Due to the attempt to replicate human reasoning, a fuzzy logic-based algorithm will be developed in this chapter. The numerical estimation results will first be transformed into degrees of change using appropriately selected and traceable membership functions. Then, a fuzzy inference system (FIS) will be developed based on the relationships between parameter changes and their associated fault modes obtained in Chapter 5.

Finally, once the performance of the entire FDD methodology is validated with the HW55S turnout, it will be validated with a different HW2000 turnout system as well. This aims to demonstrate that the methodology is not directly transferrable to different turnout systems, as it always requires some degree of re-training first. However, due to the inherent interpretability and traceability of the proposed methodology, the re-training procedures can be simplified into cross-training instead that does not require comprehensive training dataset from the HW2000 system. So, the FDD methodology developed from a parameter estimation technique and fuzzy logic-based classification is proven to be more feasibly practical for railway turnout applications.

Chapter 7: Conclusions and Future Work

This final chapter presents the conclusion of this thesis. Future works that may improve and expand on this methodology will also be recommended.

Following this thesis structure, the objectives outlined in Section 1.3 will be accomplished in the following chapters, as shown in Figure 1-3.

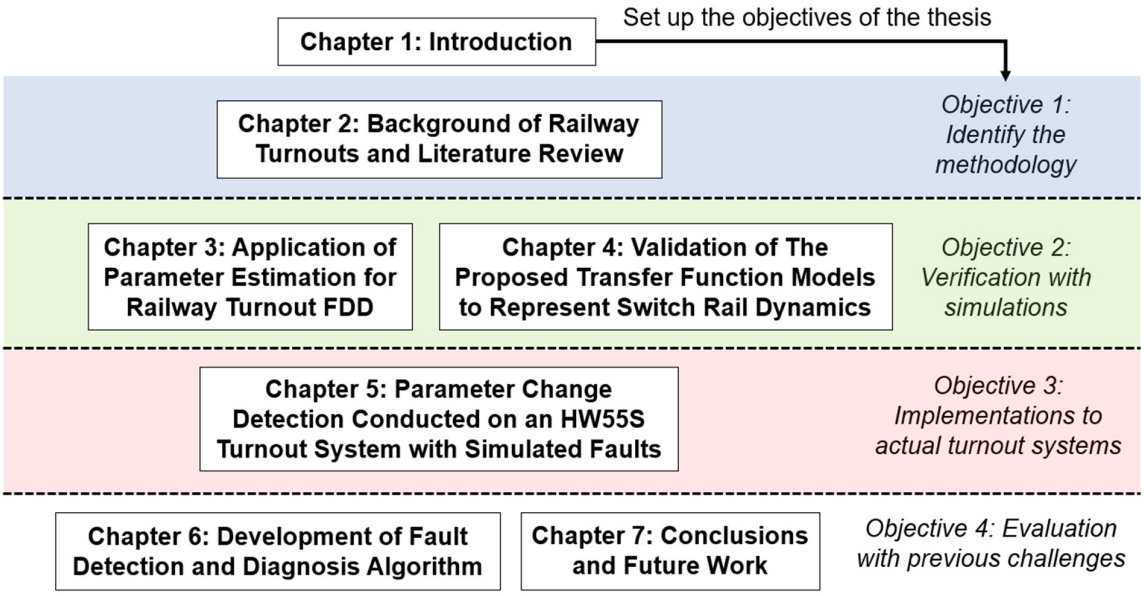


Figure 1-3 Thesis structure and chapters with their associated objectives

CHAPTER 2 BACKGROUND OF RAILWAY TURNOUTS AND LITERATURE REVIEW

This chapter provides background knowledge of railway turnouts, including their required functions, variations, and common fault modes. Then, a literature review of previous FDD developments for railway turnouts is presented. They are categorised and assessed to identify potential research gaps in this area. After the evaluation, the methodology to be developed in this PhD thesis is proposed.

2.1. Background of railway turnout systems

The first half of this chapter provides relevant background on railway turnout systems. First, it illustrates the functional procedures that every turnout must follow. Then, variations in the actuation element, commonly known as a “point machine”, are described regarding their fundamental mechanisms. Due to the quantity currently in use, electro-mechanical turnouts are the main focus of this thesis (Matei et al., 2015; Dutta et al., 2020b). At the end of this section, measurements and fault modes that were mostly considered in previous research studies are elaborated. All will serve as a foundation for the subsequent FDD development stages.

2.1.1 Functional procedures of railway turnout systems

Every railway network requires turnout systems to direct rail vehicles to the other alternating routes (Silmon-Monerri and Roberts, 2006). This is accomplished by actuating the switch rails to the desired position and securely locking them during train passages (INNOTRACK, 2008a). Therefore, railway turnouts must work in unison with the signalling systems, resulting in an integration of additional functionalities other than

simple actuation (Jin et al., 2015; Parkinson, 2019, Bradbeer, 2022). Figure 2-1 depicts a diagram of the dynamic function of turnouts in relation to man-made inputs.

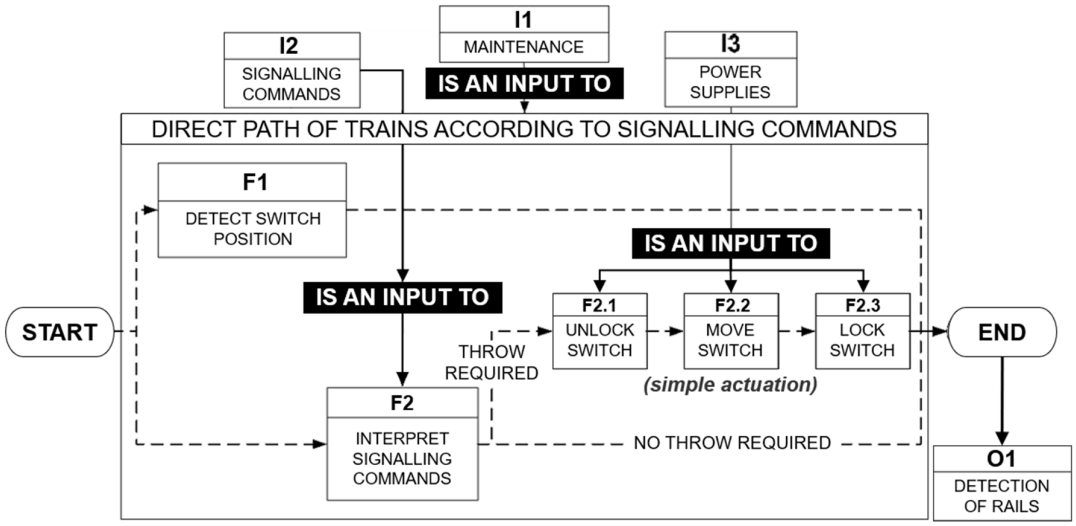


Figure 2-1 High-level diagram of the turnout's dynamic "direct path of trains according to signalling command" function (INNOTRACK, 2008a)

The performance of this function is dictated by previous maintenance, signalling commands, and power supplies. While performing this function, turnouts must be able to detect the current position of switch rails as well. Based on Figure 2-1, this dynamic function of railway turnouts can be elaborated as follows:

(1) First, before receiving a signalling command, the turnout must always be able to identify the current resting positions of the switch rails.

(2) Once the route-setting is completed, a signalling command will be sent to the turnout, which will be interpreted and compared with the current resting position to determine whether a throw is necessary.

(3) If so, the corresponding location case (LOC) will energise the turnout to start disengaging its locking mechanism, initiating the unlocking process.

(4) When the turnout has been unlocked, the driving mechanism will be engaged with the switch rails shortly after. As a consequence, the turnout can initiate the moving phase (Matei et al., 2015).

(5) Once both switch rails reach their desired positions, the locking mechanism will be re-engaged to secure them in place (Márquez, Tercero, and Schmid, 2007; How, 2022a). This mechanism is commonly known as “facing point lock” (FPL), as trains coming from the facing direction pose a greater risk.

(6) Then, two detection rods connected to each rail will interact with the detection mechanism. It will transmit the new resting position information back to the signallers. If they have received the correct update, the completion of the throw will be confirmed, and the route will be completely set for the train.

(7) After the switch rails have been successfully ***unlocked, moved, locked,*** and ***detected***, the train can proceed through the turnout as requested in Step (2).

Every turnout must successfully perform all these tasks to ensure no service disruption. Although these functional procedures are required for every turnout, the details of how it performs will differ depending on the type of point machine (Oyebande and Renfrew, 2007; Hamadache et al., 2019a). To emphasise this statement, the next section provides a brief overview of railway point machine types.

2.1.2 Common types of railway point machines

Generally, railway point machines can be classified into hand-operated and powered types. The former uses a system of rods and levers to move the switch rails from a nearby signal box. Although they are simple and inexpensive, their physical limitations and exposure to harsh environments result in their unreliability (Bemment et al., 2018). Since they also constitute less than 10% of the UK turnout population, this thesis will now only focus on powered point machines (How, 2022a).

For every powered point machine, when the signalling system has decided that a throw is requested, the associated LOC will energise the point machine. The input electrical energy will be eventually converted into mechanical work for the switch rail’s

actuation. The intermediate energy conversion processes are typically used to identify the type of powered point machines (How, 2022b; Hamadache et al., 2019a; Hu et al., 2022). There are primarily three sub-categories, electro-pneumatic, electro-hydraulic, and electro-mechanical, an overview of which follows.

Electro-pneumatic point machines are equipped with a compressor to generate compressed air. Then, the control valves will regulate and direct the airflow in a specific direction according to the command (Roberts et al., 2002). Electro-hydraulic variants operate similarly, except they use a pump motor to supply the hydraulic oil for energy transfer instead (Vileiniskis et al., 2016). One of the main drawbacks of both types is a leakage within the closed system, which deteriorates point machine performance and requires frequent inspection. Nevertheless, there are still many clamplock systems (i.e. a variation of electro-hydraulic point machines) currently in use on UK railways today, with an estimated population of 7,500 in 2021 (Jackson and Spurlock, 2021).

Instead of using fluid dynamics for energy transfer, electro-mechanical point machines utilise a direct transmission system. It consists of multiple gear stages that link the motor shaft to the driving mechanism. At the final stage, the rotating torque will be converted into an actuating force for the switch rails. Various mechanisms may be used for this purpose, including a ball-screw (SIEMENS, 2014), a lead-screw (IAD Rail Systems, 2001), or a drive bearing in an escapement (British Railways Board, 1992).

In the UK, several electro-mechanical point machine variants are used, which can be briefly outlined as follows (Jackson and Spurlock, 2021):

- M63 machines used at 750 turnouts have been deemed obsolete and are being replaced by newer technologies (Zhou et al., 2002; INNOTRACK, 2008a);
- HW variants dominate the electro-mechanical population with approximately 7,000 installations. Their first implementation dates back to the 1950s;

- High-performance Switch System (HPSS) is the latest validated technology that has just been installed on 648 turnouts (Jackson, 2023).

Due to the large quantity of electro-mechanical point machines currently in use, the research field related to them is more established than that of the others. This argument is supported by recent review papers from various countries that elaborated on several research articles related to electro-mechanical turnouts, including the UK (Hamadache et al., 2019a), Indonesia (Pratama and Hidayat, 2022), China (Hu et al., 2023), and Brazil (Narges, Ahmad, and Fereydoun, 2022).

As a result, this thesis will also focus on electro-mechanical turnouts, which was defined previously as Scope 1 in Section 1.3. The following section describes their operational behaviour, which can be observed from various available measurements.

2.1.3 Operational behaviour of electro-mechanical turnout systems

There are many measurements that can reflect the operational behaviour of electro-mechanical turnout systems (S-CODE, 2019). An example array of 16 sensors proposed by Zhou et al. (2002b) to monitor an M63 turnout illustrates the number of such availabilities, as shown in Figure 2-2.

Most of the previous FDD developments for railway turnouts focused on voltage, current, actuating force (measured by a load cell), and displacement of the switch rails as they corresponded well to the turnout's dynamic function. Figure 2-3 supports this statement by showing the times each measurement was analysed in previous turnout FDD research (Márquez, Roberts, and Tobias, 2010; Hamadache et al., 2019a).

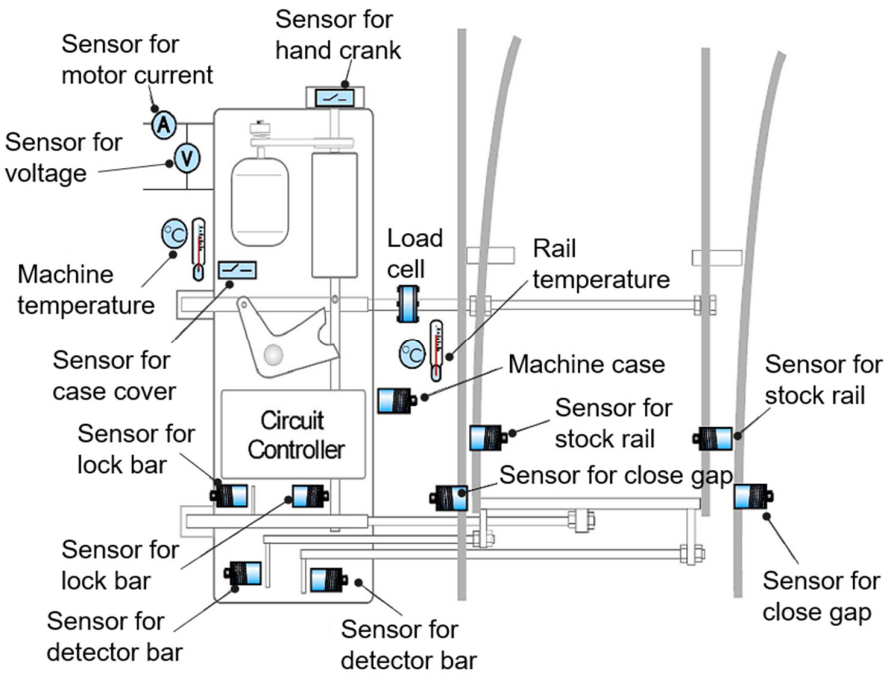


Figure 2-2 Array of 16 sensors used to monitor an M63 turnout system (Zhou et al., 2002b)

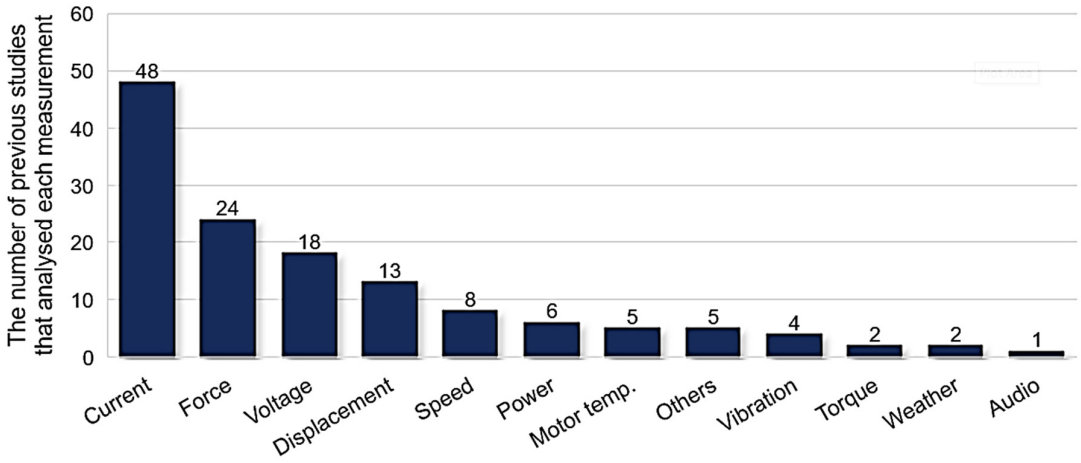


Figure 2-3 Times each measurement was analysed in previous railway turnout FDD developments (Márquez, Roberts, and Tobias, 2010; Hamadache et al., 2019a)

All four measurements (i.e. current, force, voltage, and switch rail displacement) rank among the top five. Meanwhile, “speed”, which can be measured either from the switch rails or the motor shaft, has become more prevalent in more recent studies.

The actuating force is ranked second because it is directly related to the interaction between the turnout’s driving mechanism and the switch rails (Silmon and

Roberts, 2010; S-CODE, 2019). Thus, it is normally regarded as the most informative measurement compared to the others (Atamuradov et al., 2019; Panchenko et al., 2019). However, using a load cell for the force measurement is expensive and intrusive to the turnout's function (Camci et al., 2016). As it has to be fitted into the driving mechanism, the load cell can become an additional source of turnout failures as well (Asada and Roberts, 2013). Due to such shortcomings, force measurements were still considered impractical for railway turnout FDD (Jin et al., 2015; Letot et al., 2015).

Motor current analysis is found to be commonly implemented as the current is also strongly correlated to both the actuating force and the motor torque. Furthermore, current measurement is more practical for turnout applications because it is more cost-effective, non-intrusive, and straightforward (Camci et al., 2016; Cui, Tang, and Ou, 2019; Guo et al., 2020a; Chen et al., 2021). Examples of current data obtained from fault-free throws of several point machines are shown in Figure 2-4.

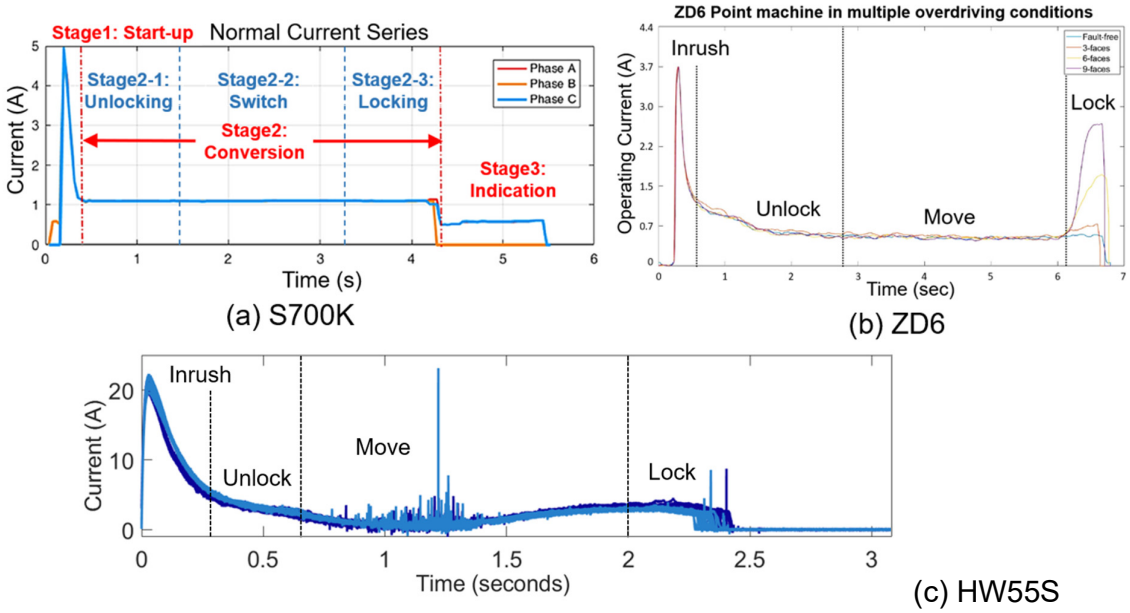


Figure 2-4 Examples of segmented current waveforms (a) S700K; it can be segmented into either three (Ou, Xue, and Cui, 2019) or five phases (Zhao et al., 2018), (b) ZD6 (Zhao et al., 2017; Kumpao, 2019), and (c) HW55S; it is available as a test subject in this thesis

It can be seen that although every turnout performs the same function, different point machine designs will result in different current waveforms to some extent. These currents can also be divided into multiple phases corresponding to each step of the turnout's dynamic function to facilitate better FDD analysis. This process is commonly known as ***"segmentation"***. Figure 2-4(c) illustrates the current measurements obtained from an HW55S point machine, which is available as a test subject in this thesis. Its operating phases during a throw, after the segmentation, can be described as follows:

(1) ***Inrush***: When the motor is energised, it must overcome the combined inertia of its shaft, the transmission system inside the point machine, and the standstill switch rails (Soares et al., 2021). Thus, the maximum torque (and motor current) of the entire throw can be observed in this phase.

(2) ***Unlock***: After the transmission system inside the point machine starts to move, its first objective is to disengage the locking mechanism and unlock the switch rails. This manoeuvre allows the rails to be moved freely afterwards.

- A change in the current decreasing rate can be observed in this phase, as the switch rails will try to return to their equilibrium while being held by the locking mechanism. This results in the first mechanical resistance.

(3) ***Move***: When the switch rails have been unlocked, their internal stiffness will restore them to their equilibrium position, aiding the point machine. Once the driving mechanism is fully engaged with the switch rails, their actuation to the other position will be initiated (Bunnett, 2015).

- As the switch rails are moved further from their equilibrium position once again, the rail's actuation will be more resistant over time (Oyebande and Renfrew, 2007). This results in the second mechanical resistance.

(4) ***Lock***: When they reach the other end position, the motor must continue to operate briefly to re-engage the locking mechanism.

- During this time, the motor current may increase again as the closing switch rail is being pressed against the stock rail. This increase should not be significant if the turnout system is correctly aligned and adjusted.

After a successful lock, the motor contacts will be isolated, stopping the current from the LOC. Then, an internal snubber mechanism will reduce the circulating current and the residual motor shaft speed to zero, completing the throw.

The detection phase does not appear in the HW55S current waveforms, as this function is generally performed by a separate set of contacts and a separate power supply (British Railways Board, 1992). Therefore, motor current analysis alone cannot inform the detection performance of such turnout systems.

S700K and ZD6J point machines are exceptions. Their detection phase appears in the current waveforms after the locking phase, see “Stage 3: indication” in Figure 2-4(a). This is because they utilise a separate “indicating” circuit, which will only be connected after a successful lock to indicate the rail’s end position (Ou and Xue, 2018; Kun, Tao, and Cheng, 2021). Since it does not apply to most electro-mechanical point machines, the detection performance will not be further discussed in detail.

Other measurements have also been employed to observe turnout operational behaviour. An example of speed and power analyses of the Alstom DK FBANE MET point machine is demonstrated in Figure 2-5 (Jin et al., 2015; Letot et al., 2015).

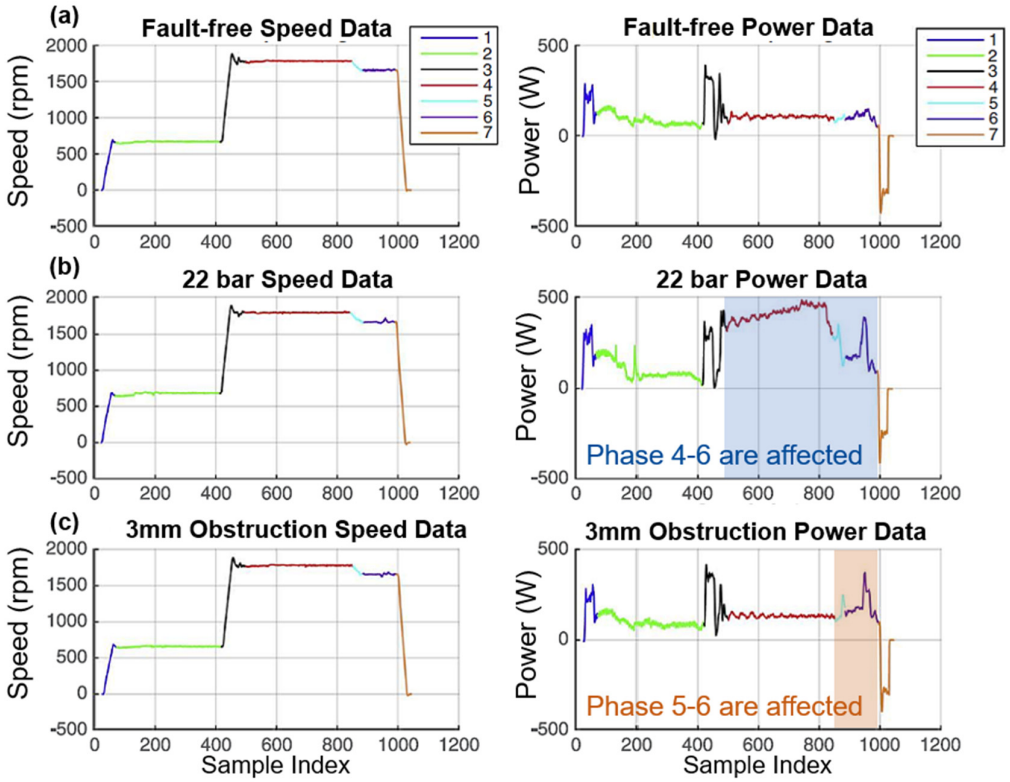


Figure 2-5 Segmentation of motor speed and power measurements from Alstom DK FBANE MET point machines into seven phases and fault simulation results (Shi, Liu, and Lee, 2018)

A thorough understanding of point machine mechanisms allows researchers to perform segmentation and develop more effective FDD algorithms, as different fault modes affect turnout systems differently (Bemment et al., 2018; Guzman et al., 2022). Figure 2-5(b) and (c) demonstrate examples of measurements obtained during fault simulations when an extra hydraulic device was inserted to simulate an increase in friction in the opposite direction and when a 3 mm obstruction was placed between the closing gap, respectively. Compared to their fault-free measurements in Figure 2-5(a), changes only appear in some particular phases of the power data.

Specific knowledge about Alstom DK FBANE MET point machines will allow the maintainers to capture these unique characteristics and diagnose faults correctly. As the significance of system knowledge is demonstrated here, the following section will provide insight into the common fault modes of electro-mechanical turnout systems.

2.1.4 Common fault modes of electro-mechanical point machines

Fault modes that commonly occur in electro-mechanical turnout systems are described in this section. Their characteristics and effects on turnouts are outlined, along with fault simulation processes previously used for data generation.

The objectives of FDD systems in turnout applications are to detect early ***fault*** developments and rectify them before they escalate to become ***failures*** (Kim et al., 2016; Zoljic-Beglerovic, 2018). The standardised definitions of “fault” and “failure” provided by The International Federation of Automatic Control (IFAC) Technical Committee on Fault Detection, Supervision, and Safety to Technical Processes will be consistently used throughout this thesis (IFAC, 2019).

- ***Fault***: An event in which a turnout behaves differently from its fault-free state to some extent. However, it must still accomplish all procedures outlined in Section 2.1.1 (i.e. unlock, move, lock, and detect). It usually appears as increases or fluctuations in the actuating force (and motor current) or a longer throw time. As a result, they are essentially the results of system *degradation*.
- ***Failure***: An event in which a turnout fails to perform at least one of its required procedures. Signallers will be immediately notified because it will result in an unsuccessful route-setting, leading to *service disruptions* and *delays*.

They can be further classified based on their temporal characteristics as follows:

- *Abrupt:* A sudden fault or failure occurs without any prior indication. Its effects will remain in the system until corrective maintenance is carried out.
- *Intermittent*: A fault or failure occurs irregularly and repeatedly without prior indication. However, its effects often randomly disappear after a while, which could be before the arrival of the maintenance staff (Asada, 2013).

- *Incipient*: A degradation that gradually develops inside a system over time. It deteriorates the system from being fault-free to a fail state. Therefore, it usually describes the progression of faults.
    - Characteristics of incipient faults allow FDD methodologies to perform their intended functions more effectively by detecting the early stages of fault developments, which do not appear in the other two types.

All these three temporal characteristics are illustrated in Figure 2-6.

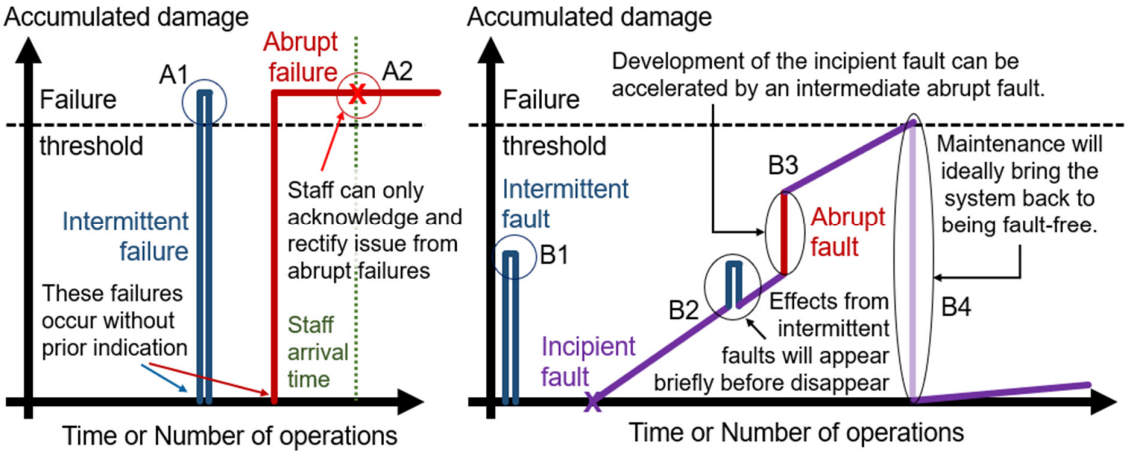


Figure 2-6 Illustrations of faults and failures temporal characteristics (Author, 2023)

The left-hand side of Figure 2-6 illustrates the effects of intermittent and abrupt failures. Neither shows any signs of prior indication, so the staff cannot prevent or anticipate their occurrences. Due to the nature of intermittent failures (A1, in Figure 2-6), their effects may disappear before the staff arrive. Since the turnout may have returned to its fault-free state already, the staff can only record in their databases as "Tested OK on Arrival" or "No Fault Found" (Pecht and Jaai, 2010). Meanwhile, the effects of abrupt failures (A2) are observable and rectifiable.

The effects of the three fault types are shown on the right-hand side of Figure 2-6. Here, the effects of intermittent faults (B1) will not impact the system sufficiently and cause it to fail. They may randomly reappear during the development of incipient

faults (B2). Meanwhile, when incipient faults have initiated, they will gradually become more severe over time. The development of these faults could be accelerated if an abrupt fault happens in the meantime (B3). Once the system has failed and corrective maintenance has been conducted (B4), it should ideally return to being fault-free.

Due to the unpredictable of abrupt and intermittent faults, most previous FDD developments for turnouts did not focus on their detection (Zhou et al., 2002b; Asada and Roberts, 2013). More emphasis was given to incipient faults instead since their emergence and development were more deterministic and observable than those of the others. Successful detection of early incipient faults would also facilitate effective fault diagnosis, prognosis, and a closer realisation of CBM (Chen et al., 2020).

Due to the advantageous contributions it provides, this thesis will only focus on incipient fault diagnosis. To select which fault modes to consider, Figure 2-7 indicates the number of acknowledgements of common fault modes found in the previous 35 related publications. It also includes the number of publications that mentioned their fault simulation processes for data generation purposes.

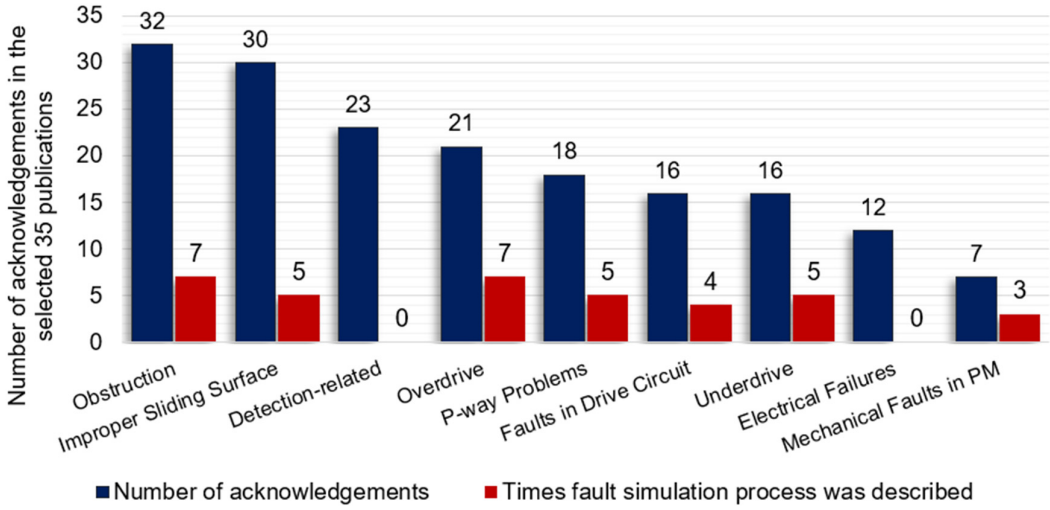


Figure 2-7 Number of acknowledgements of each common fault mode of electro-mechanical turnout systems and times when their simulation process was described

Different authors would normally use different terms to describe the same fault mode. For consistency in presentation, Table 2-1 provides some key terminologies that appeared in the selected research studies based on the fault categorisation used in Figure 2-7. A full list of them can be consulted in Appendix 1.

Table 2-1 Some key terms used in the 35 selected publications that related to the common fault modes being categorised as in Figure 2-7

Common fault modes (corresponding with the graphs in Figure 2-7)	**Terms found in the selected publications**
Obstruction	Changing the fulcrum point; Ballast/ice obstruction; Switch machine idle; Jamming of the throw bar.
Improper Sliding Surface	Resistance increases due to temperature; Dry/ contamination/deterioration/frozen/defective chairs.
Detection-related	Switch cannot detect; Indicating rod block in the gap; Indicating circuit diode damage; Microswitch defective.
Overdrive	Switch rail rebounds while locking (with noise and hunting); Hard to release; Adjuster bolts misalignment.
P-way Problems	Misalignment of stock rails; Movement of the sleepers and rails; Broken stretcher bar; Incorrect installation.
Faults in Drive Circuit	Increase in winding resistance; Low tension on motor brush; Reduced input voltage; Bad/poor contact.
Underdrive	Lack/out of adjustment of lock/drive bars; Clearance between rails; Fail to lock; Adjuster bolts misalignment.
Electrical Failures	Major electrical system failure; Supply interruption; Drive circuit defective (from fuse, relay, contacts).
Mechanical Faults in Point Machine	Unstable work of the friction clutch; Slit up of point machine’s bearings; Gear misalignments.

Obviously, there are more than 35 publications in this research field. However, most of them only aimed to detect faults or failures without identifying their nature and cause. Therefore, there was no information regarding fault modes in those studies. As a result, Figure 2-7 does not reflect the accurate significance of each fault mode but only provides a general perception based on an academic standpoint.

Only five fault modes in Figure 2-7 are discussed in greater detail in the next sections. The justifications for not considering the other four are:

- None of the 35 selected papers provided the simulation process for "detection-related" faults. Those that investigated detection-related faults focused only on point machines with an indicating circuit for detection, where such faults cause noticeable differences in current measurements. However, the HW55S system — available as the research object in this thesis — employs a different detection mechanism, such that the detection-related faults no longer influence current measurements. Thus, the effects of this fault mode on turnout dynamics are neither observable nor reliably replicable through fault simulations.
- “P-way problems” faults are not included in the scope of this thesis, as stated in Scope 2 in Section 1.3. Moreover, it is difficult to simulate them on the available slab track testbed in a laboratory.
- “Electrical failures” will not be considered as they are not *faults*, such that throws cannot even be completed in these situations.
- “Mechanical faults in PM” were of least concern, as the point machine itself is usually self-contained and well-protected by its tight and weatherproof cover.

2.1.4.1 Obstruction fault

An obstruction fault is *when a foreign object is in the closing gap between the switch rail and the stock rail*. Due to its universally simple mechanism and frequent occurrences, it is commonly investigated (see Figure 2-7) and conveniently simulated.

The most common obstruction in ballasted track railways was ballast and its debris. A thin layer of “soft” obstacle can also be formed in the closing gap, such as compressed leaves, build-up of ice (Lee et al., 2016), old lubricant on the slide chairs,

and even layers of snow (Wang et al., 2017). Other unusual objects, like bottles, litter, and a forgotten wrench, may also be mistakenly left by maintenance staff.

When this type of fault occurs, the actuating force and motor current will be increased at the end of the throw as the motor tries to compress the obstruction (Matei et al., 2015; Kun, Tao, and Cheng, 2021). Furthermore, it also leaves a clearance gap between the two rails, which equals the size of the compressed obstacle.

In the case of faults, the compressed obstruction is not larger than the tolerance of the locking mechanism, which is normally between 3 and 5 mm, depending on the point machine design (Zhou et al., 2001), as shown in Figure 2-8(a). The throw still will be completed in these situations, as the locking mechanism can still be re-engaged.

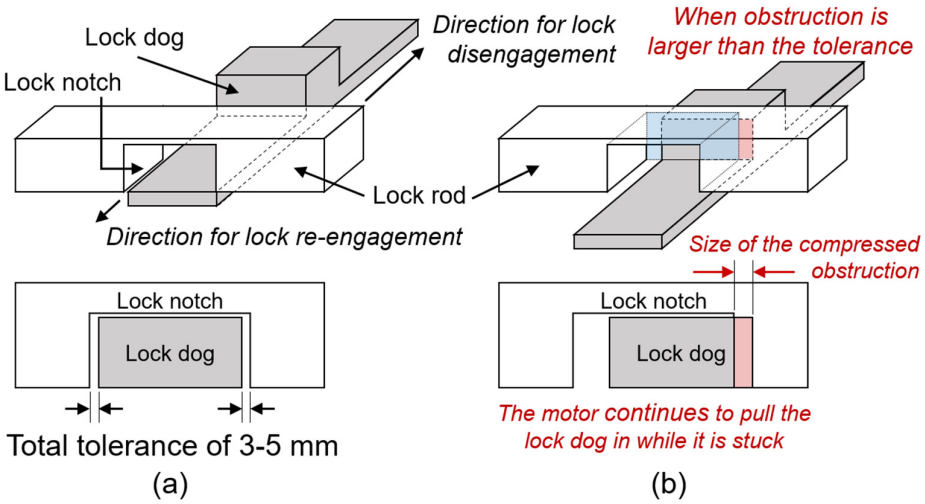


Figure 2-8 General structure of a railway turnout locking mechanism operating in (a) fault-free states and (b) during obstruction failures (modified from Igarashi and Siomi, 2006)

Once the obstruction size is larger than the tolerance, the motor contacts will not be isolated, resulting in an uninterrupted supply. This happens because the lock notch and the lock dog are not properly aligned and stuck to one another, as shown in Figure 2-8(b). As the throw is incomplete and the motor is still energised continuously, it will keep pulling the lock dog in, leading to a large mechanical resistance.

This fault simulation is simple because any arbitrary obstruction can be inserted between the closing gap, e.g. ice and ballast (Lee et al., 2016). Meanwhile, simulation of multiple severity levels is also possible by placing shims of different thicknesses at different locations (Márquez, Schmid, and Collado, 2003a; Letot et al., 2015).

2.1.4.2 Improper sliding surface

An improper sliding surface fault is when *a higher level of resistance is present during the movement of the switch rails due to abnormal conditions of the slide chair's surface*. Previous failure databases demonstrate that this fault happens frequently, which could be due to exposure of the turnout to outdoor environments.

Slide chairs are normally used to reduce mechanical resistance while the switch rails are actuated. When they cannot ease the actuation as intended, this kind of fault may have been developed in the turnout (Dutta et al., 2020b; Sun et al., 2020).

Considering faults imposing consistent effects, which is when the slide chair's surfaces are uniformly deteriorated, this condition will consistently increase friction and the required actuating force (and motor current) during the moving phase. It could result from a temperature change, which causes the lubrication to be less effective although it has been applied recently and evenly (Jin et al., 2015; Zhou et al., 2016), or when heavy rainfall washes away all of the lubricants (Wang et al., 2017).

The presence of rust patches on slide chairs (Guzman et al., 2022) or defective roller bearings (Matei et al., 2015) can create inconsistent effects as well. Meanwhile, changes in a turnout's static characteristics, including the movement of bearers and rails, maintenance mistakes, tampering, and movement of the ballast format, also contribute to increased resistance (Ou and Xue, 2018; S-CODE, 2019).

Incipient faults of this mode can develop in two ways. First, when the moving resistance increases, the throw time will be longer. Once the throw time has exceeded the setting of the timer relay, the power supply will be isolated during a throw, resulting in a failure. Second, contamination on the surfaces can build up at the closing gap every time the turnout is operated. It will become an obstruction failure when the layer of compressed contamination is larger than the locking mechanism's tolerance.

Previous studies focused on the simulation of incipient faults, which could be designed to mimic the gradual degradation. Figure 2-9 demonstrates two approaches previously employed, with Figure 2-9(a) as an example of a failure contamination scenario. Atamuradov et al. (2019) poured a thick layer of sand onto a number of slide chairs incrementally until the turnout could not complete a throw within the time required; see Figure 2-9(b). In Figure 2-9(c), Shi, Liu, and Lee (2018) installed an additional hydraulic device to generate a resistive force. They gradually increased the hydraulic pressure to a maximum of 22 bars, which caused the turnout to be idle.

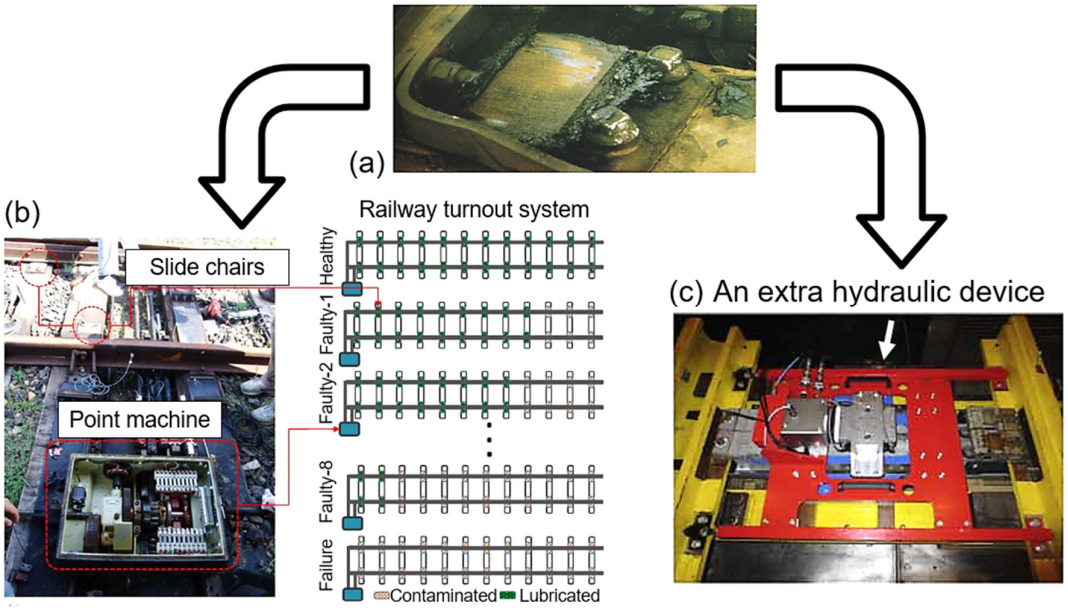


Figure 2-9 Examples of incipient improper sliding surface fault (a) a failure condition (Hassankiadeh, 2011), (b) simulated by increasing the number of contaminated slide chairs (Atamuradov et al., 2019), and (c) using an extra hydraulic device (Shi, Liu, and Lee, 2018)

It was also discovered that effects from this kind of fault are turnout- and direction-specific (Márquez, Schmid, and Collado, 2003b). Some scenarios may pose severe effects, while others may have none. In accordance to this, Márquez, Weston, and Roberts (2007) found that the latter circumstances are likely to occur if the entire length of switch rails does not make firm contact with the slide chair's surface.

2.1.4.3 Overdriving fault

An overdriving fault is caused by *an incorrect driving mechanism. It results in the switch rails being continuously pressed against the stock rail, even after the gap has completely closed*. Therefore, faulty throws can still be completed, but mechanical linkages will experience excessive built-up stress.

Electro-mechanical point machines normally perform unlocking, moving, and locking tasks sequentially (Bemment et al., 2018). As the unlocking process must be completed before the switch rails actuation, an "adjuster" mechanism is required to delay the moving phase until the unlocking process has been completed (Oyebande and Renfrew, 2007; INNOTRACK, 2008b; Matei et al., 2015). Figure 2-10 illustrates this "adjuster" mechanism used in HW55S point machines, assuming the switch rails rest in the "normal" position. It consists of a drive lug attached to the drive stretcher bar while being trapped between two sets of adjuster lock nuts.

The use of this mechanism will create a "lost-motion" distance, which equals a part of the drive rod stroke while the switch rails are still at rest. For overdrive faults, the lock nuts would be located closer to the drive lug than they are nominally designed (see "overdrive adjustments" in Figure 2-10 and 2-11).

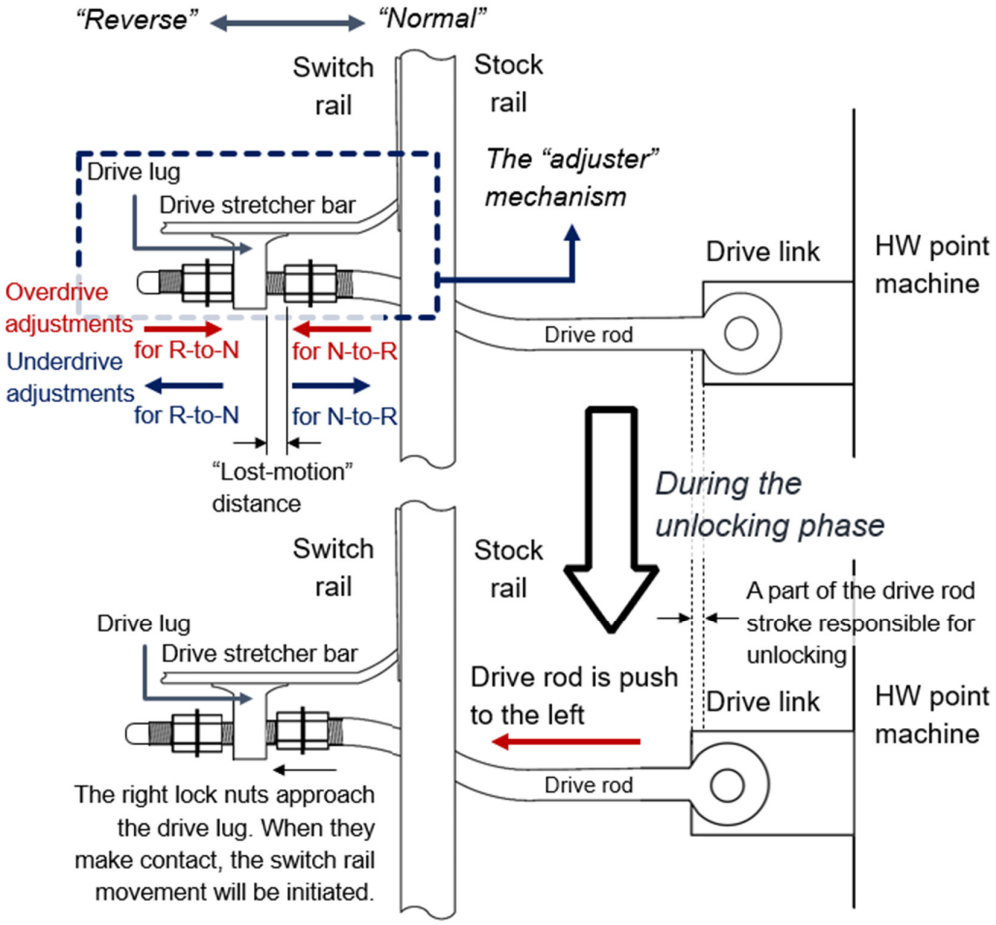


Figure 2-10 The driving and the "adjuster" mechanism that provides a "lost-motion" distance for HW55S point machines (modified from INNOTRACK, 2008b)

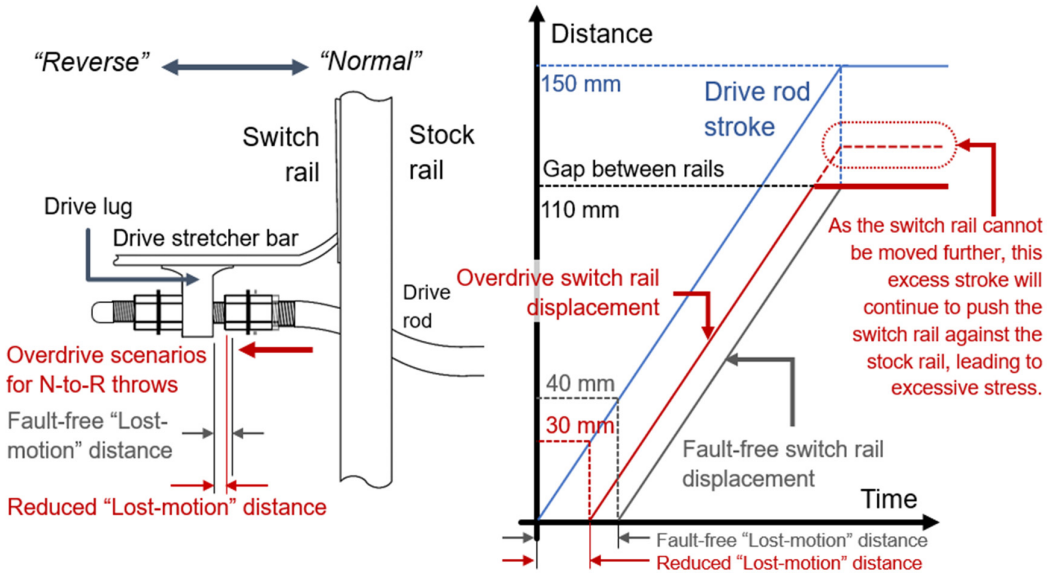


Figure 2-11 Overdrive fault simulations with the trajectory of the switch rails (Author, 2023)

This will effectively reduce the lost-motion distance and time delay, which initiate the moving phase sooner than intended and extend the actuating duration. As a result, the switch rails will be pressed against the stock rail, even after the gap has been fully closed. The motor current will also be increased during the locking phase if the motor can overcome the switch rail's stiffness (Chen et al., 2021; Guzman et al., 2022).

When overdrive faults happen, the unlocking process of the following throw (i.e. in the opposite direction) will be more resistive (Zhou et al., 2016). Disengaging the locking mechanism will now require a much higher pulling force. This circumstance is commonly referred to as "unlock difficulty" (Cheng and Zhao, 2015; Zhao et al., 2017).

The simulation of overdrive faults can be conducted by moving the lock nuts, which can also be done at multiple severity levels (Asada and Roberts, 2013; Chen et al., 2020). In reality, it may be caused by traffic vibration, mistakes during maintenance, or changes in ambient climates (Iwasawa et al., 2015). Due to the delicacy of these adjustments, according to a report by Network Rail in 2008, overdrive scenarios are responsible for 9% of all turnout failures on UK railways (Silmon, 2009).

As a throw can still be completed in the presence of overdrive faults, they should be detected as early as possible. Inability to do so will lead to a build-up of stress and fatigue, which can cause fractures in stretcher bars, fasteners, and brackets. These may happen during train passages and cause derailments (Ball, 2010). Being one of many causes of the Potter Bars accident, it could have been prevented if an effective FDD system had been implemented (Márquez, Roberts, and Tobias, 2010).

2.1.4.4 Underdriving fault

An underdriving fault is an event caused by *incorrect adjustments in the driving mechanism as well, just like an overdriving fault. However, it creates opposing effects, where the switch rails are not adequately actuated to close the gap completely*.

When the lock nuts are located farther from the drive lug than they are nominally designed (see the "underdrive adjustments" in Figure 2-10 and 2-12), the lost-motion distance and time delay will be extended instead. The point machine will still be able to lock the switch rails if the remaining gap is smaller than the tolerance of the locking mechanism (Jin et al., 2015). When the tolerance cannot accommodate a gap left by an underdrive fault, the situation shown in Figure 2-8(b) will occur. The point machine will then be unable to lock and detect the switch rails, resulting in a failure.

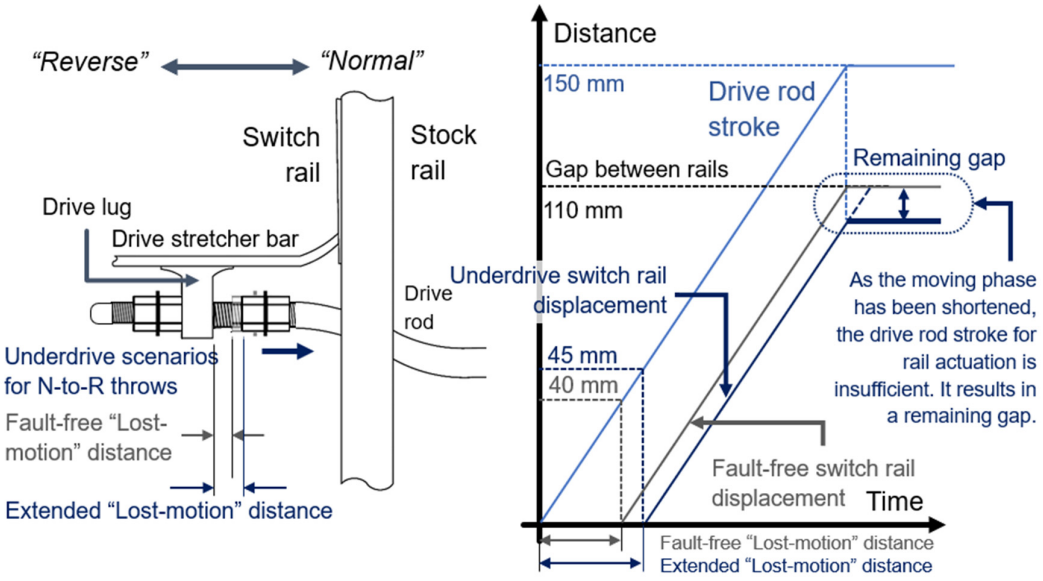


Figure 2-12 Underdrive fault simulations with the trajectory of the switch rails (Author, 2023)

The simulation of underdrive faults also involves repositioning the adjuster lock nuts. However, in this case, the increase in the actuating force and motor current will not appear during the locking phase. That is because the switch rails will not completely close the gap nor make contact with the stock rail, resulting in a clearance gap similar to the obstruction faults (Asada, Roberts, and Koseki, 2013; Matei et al., 2015).

2.1.4.5 Faults in Drive Circuits

These faults are events that are caused by *abnormal electrical behaviours in any component of the drive circuits*. The consideration starts from the circuitry inside the LOC, the cables that proceed through multiple connection boxes, enter the point machine closure, and finally connect to the internal electric motor.

First, components inside the LOC may be malfunctioning (Ou and Xue, 2018). Voltage variation caused by a defective power supply is one of the incipient electrical faults that can be reliably simulated (Oyebande and Renfrew, 2002; 2007). An input voltage that is not lower than the acceptable range will not extend the throw time beyond the setting of the timer relay, only resulting in a fault.

Second, the condition of cables and electrical connections will affect the overall circuit resistance (Cui, Tang, and Ou, 2019; Ou et al., 2020). These faults happen because the drive circuit can be up to kilometres in length, which is all exposed to unforgiving railway environments (Guzman et al., 2018). Damage to them can cause electrical faults that eventually lead to open-circuit failures.

Finally, there may be problems with the electric motor itself. Brushed DC motors are prone to worn and defaced brush conditions, while AC motors are immune to them (Márquez, Schmid, and Collado, 2003a). Other problems include damaged rotors from improper loading conditions and misalignment (Dutta et al., 2020b) and varying the winding resistance due to inter-turn short-circuits (Dixon and Pike, 2002).

According to previous failure databases, electrical faults do not occur as often as the other types. This led to a conclusion by Hassankiadeh (2011) and Hamadache et al. (2019a), who reported that there had not yet been a thorough investigation of

electrical faults in this research field. Therefore, the proposed FDD methodology in this PhD thesis will also cover incipient voltage variation faults.

2.1.5 Conclusions

This section covered the background knowledge of railway turnout systems. It began with a description of the functional procedures that every turnout must follow. An overview of various point machines, particularly the powered variants, has been provided. Due to the large quantity currently in use, electro-mechanical turnouts were selected as the main subject of this thesis. Several measurements that possess the ability to reflect turnout performance have also been described. The segmentation process was then introduced to emphasise on the effects of different fault modes on particular phases. Finally, fault modes commonly affecting electro-mechanical turnout systems were investigated based on previous studies. Such background knowledge will be of paramount importance for the subsequent FDD developments.

2.2 Literature review of railway turnout FDD developments

In this section, previous FDD developments for railway turnouts are evaluated to identify research gaps in this area. A number of related articles have been published since the early 2000s, when data acquisition technologies had become more readily available to the industry (Zhou et al., 2002a; Shaw, 2008).

Due to false alarms and missed failures from the early thresholding-based FDD approaches, many scholars have proposed more advanced data analysis techniques that provide superior accuracy. If their implementation is successful, they will reduce the unnecessary workload, as the number of turnout measurements to be analysed further could be decreased by automatically ignoring the fault-free ones.

Regardless of various techniques being implemented in any FDD algorithm, their procedures always consist of multiple transformations, from raw measurements to identifying the class of current health state, as shown in Figure 2-13.

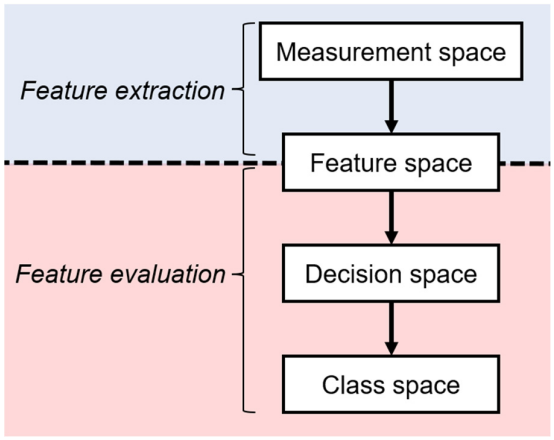


Figure 2-13 High-level transformation stages in every FDD algorithm (Wilhelm et al., 2021)

The first stage will transform raw system measurements in a high-dimensional space into representative and lower-dimensional **features**. This process is commonly known as *"feature extraction"*. Distributions of features extracted from known fault-free and faulty throws are required to formulate the decision boundaries inside the feature hyperspace. The process of comparing features from a test throw to those boundaries is called *"feature evaluation"*, which will produce the final diagnostic result.

In this thesis, previous FDD developments for railway turnouts are categorised into "data-driven" and "model-based" methodologies based on **feature extraction**. Regardless of the terms used in this categorisation, both of them still require training datasets and system knowledge to some extent. Such overlaps have resulted in these terminologies being defined differently by different authors (Venkatasubramanian, Rengaswamy, and Kavuri, 2003a; Zhang and Jiang, 2008; Gao, Cecati, and Ding, 2015; Jimenez et al., 2020). As a result, definitions of these two categories are given below and will be consistently referred to throughout this thesis.

- ***Data-driven methods***: Features are selected and calculated from pre-processed system measurements directly, without the consideration of relationships between the system's inputs and outputs. So, these methods do not rely on the underlying physics and features from inputs and outputs are usually considered independently.
- ***Model-based methods***: Features are calculated from a mathematical model which aims to capture relationships between the system's inputs and outputs. Therefore, the extracted features will be dependent on multiple measurements.

It should be noted that the aim of this section is not to provide a comprehensive review of all previous FDD developments. It will only highlight their characteristics and limitations when they are explicitly applied to railway turnout systems, which facilitates the identification of the potential research gaps.

2.2.1 Data-driven methodologies

The term "data" refers to measurements collected from the monitored system, including both its inputs and outputs. Directly extracting features from these data and analysing them is attractive when their patterns are recognisable (Jimenez et al., 2020) and when precise mathematical models governing the system's dynamics are difficult to obtain (Dowdeswell, Sinha, and MacDonell, 2020).

The available data are normally divided into training and testing datasets. The former is mandatory to define decision boundaries. Therefore, distributions of features extracted from the training dataset should ideally have low variance and be well-separated between different health states to avoid underfitting (Pecht and Jaai, 2010). Then, the testing dataset will assess the approach's performance and generalisability.

In this thesis, the author further divides data-driven FDD methodologies into four sub-categories based on how the features originated, namely expert knowledge, similarity analysis, statistical features, and deep learning-based approaches.

2.2.1.1 Expert knowledge-based approaches

Experts can help identify the most effective features that indicate a system's current health state. Based on the extent of their knowledge and experience, they may be able to determine the optimal thresholds used in the evaluation process as well (Jimenez et al., 2020). Therefore, the algorithm's objective is to possess the same level of expertise as if it is another expert in that particular system.

This concept has been implemented for railway turnouts since the early 2000s. Oyebande and Renfrew (2002; 2007) demonstrated this by evaluating measurements gathered from laboratory and in-field HW2000 turnout systems. Their team of experts suggested several features to consider, whereas the total energy used during a throw provided the best indication of fault characteristics. It was robust to signal noises and voltage reduction while corresponding well with previous maintenance records.

Márquez and his team proposed that wear-related faults on slide chairs have consistent effects on the absolute differences between the force measured from testing operations and the dynamically updated fault-free reference (Márquez, Schmid, and Collado, 2003b). After that, they proposed three criteria based on the characteristics of the absolute difference waveform for fault detection, as shown in Figure 2-14. The performance was improved further after the application of filtering algorithms, such as Fixed Interval Smoother (FIS) and Kalman Filter (KF) (Márquez and Pedregal, 2006).

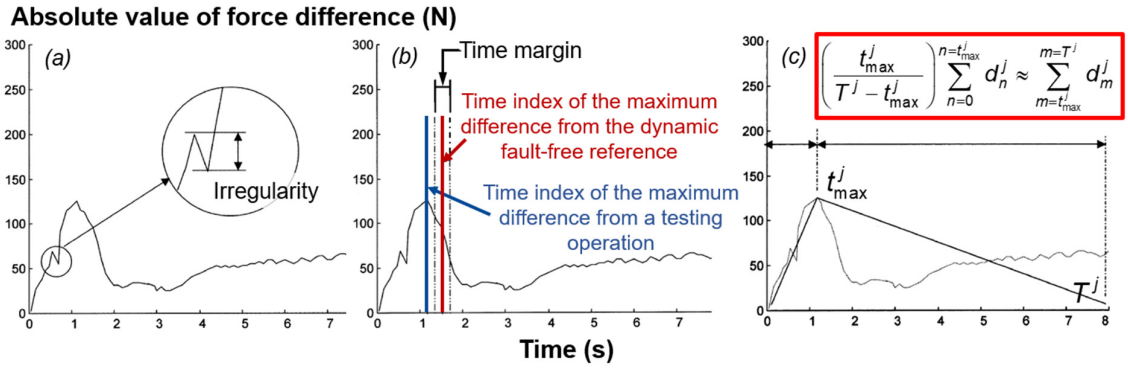


Figure 2-14 Three consecutive criteria that experts have proposed for wear-related slide chair fault detection: (a) irregularity, (b) time index of the absolute maximum force difference, and (c) symmetry of the waveform (Márquez, Schmid, and Collado, 2003b)

Different turnout systems and experts' experiences have usually resulted in varying features being proposed. For instance, Ou and Xue (2018) proposed a fault feature logic diagram based on the S700K current waveform's features, such as the existence of the indicating phase, phases that contain abnormal duration, and the duration and average current of the starting phase. Meanwhile, Guzman et al. (2022) proposed an interpretable diagnostic system for NSE turnouts by evaluating humps and flat spots in their current waveforms, throw time, and resting time between throws.

All in all, expert knowledge-based approaches are interpretable and intuitive, making their reasoning easy to understand by maintenance staff. These properties are deemed essential for safety-critical applications because their methodologies can be justified by infrastructure managers (Chen and Roberts, 2006; Bukhsh et al., 2019). Furthermore, features used in these approaches are also simple to calculate compared to those employed in more sophisticated algorithms (Zhou et al., 2016).

However, their evaluation step still relies on preset thresholds, which can be determined by the training dataset, if available. Otherwise, they are entirely based on the experts' experience. Due to more railway experts becoming retired nowadays, the existing staff may not acquire sufficient knowledge to set thresholds appropriately

(Zhao et al., 2018). Their judgement can also be prone to subjectivity and human error, resulting in poor performance (Bemment et al., 2018; Li et al., 2019). As a result, these approaches are generally not applicable beyond the experts' expertise.

2.2.1.2 Similarity analysis-based approaches

Suppose experts are not available, or their domain knowledge is insufficient for FDD purposes. In that case, waveform comparisons on a sample-by-sample basis can be conducted instead. In this section, methodologies based on similarity analysis are used to examine the differences between the reference waveform (obtained from known fault-free operations) and the measurements gathered from testing operations.

A comparison of qualitative trends between two waveforms can be regarded as ***shape analysis***. It is intuitive, similar to how humans would analyse waveforms, and provides "visual interpretability" (Sa et al., 2017a; S-CODE, 2019). This concept was demonstrated in the INNOTRACK project, where qualitative changes in the waveforms obtained from multiple fault simulations were evaluated. Then, Silmon (2009) proposed the Qualitative Trend Analysis (QTA) algorithm, which assessed the numerical values of the trend changes as well. It was discovered that only the qualitative sides of the algorithm were transferrable to other turnouts, while the quantitative aspects must be adjusted to fault-free characteristics of those other turnouts first before the algorithm can be implemented for them (Silmon and Roberts, 2010).

Despite the transferability of qualitative aspects shown in the QTA algorithm, most of the similarity analysis-based approaches still consider quantitative differences. For example, Márquez, Tercero, and Schmid (2007) applied Unobserved Component (UC) models to estimate the noise-free force measurement and compared it to the fault-free reference to compute the correlation coefficient. This criterion was evaluated via the preset threshold for fault detection. Then, applications of the Moving Average

(MA) filter were found to be more effective, as they could facilitate the detection of multiple fault modes in both directions (Márquez, Weston, and Roberts, 2007).

Their following study proposed an application of forecasting and the Harmonic Regression (HR) model, where the period of every harmonic is predicted from previous throw times of known fault-free throws (Pedregal, Márquez, and Roberts, 2009). If the variance of the differences between the HR regression result and the measured motor current exceeds the preset threshold, the algorithm will issue a fault alarm.

Due to random variations of the throw time, even from the same turnout design and working conditions, dissimilarity measures calculated from traditional approaches (such as Euclidean or city block) can produce unsatisfactory results (Huang et al., 2017; Du et al., 2019). This characteristic led to more sophisticated algorithms, such as Dynamic Time Warping (DTW) (Atamuradov et al., 2009) and Edit-Distance with Real Penalties (ERP) (Vileiniskis et al., 2016), being introduced to enable more effective comparisons between two time series of different durations.

For instance, Kim et al. (2016) employed the DTW algorithm to motor current from in-field turnout systems. The accumulated difference threshold based on expert experience was then recommended for fault detection. Furthermore, many variants of DTW algorithms have been introduced in this research field. First, the derivative DTW implements the squared difference of the approximated derivatives to calculate the dissimilarity between two time series instead of using the Euclidean-based calculation (Ou et al., 2018). Non-linear functions can be used to facilitate the analysis of non-linear dissimilarity measures between health states (Wang et al., 2023). Improvements in the computational time can be achieved by using the FastDTW algorithm (Gao et al., 2022), while performance improvements can be achieved by combining the DTW results and expert knowledge-based features (Zhao et al., 2018).

When the differences are subtle, similarity analysis of the entire time series can underperform due to redundant information (Eker, Camci, and Kumar, 2012). As a consequence, Sa et al. (2017a; 2017b) proposed an analysis based on a particular subsequence of motor current waveforms instead and extracted its shapelet feature; see Figure 2-15. According to their team of experts, the selected subsequence can reflect subtle ageing effects better than the rest of the throw. Then, the shapelet was evaluated to determine if the turnout should be replaced.

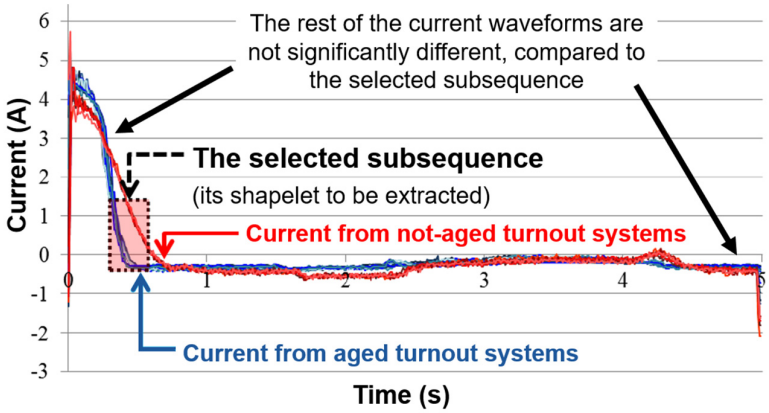


Figure 2-15 Similarity analysis based on the selected subsequence (Sa et al., 2017a)

In conclusion, these approaches focus on differences between the reference fault-free waveform and one obtained from a testing operation. Expert knowledge is not required to specify which features to consider as every sample is evaluated, resulting in no risk of information loss. Moreover, similarity analysis-based methods can provide visual interpretability to the users as well.

A more effective comparison between two time series of different durations and filtering algorithms, such as MA filters, UC models, and HR, has been introduced to improve the fault detection performance of similarity analysis-based approaches (Hu et al., 2022). Segmentation has also proven to be a huge benefit because it helps confer fault diagnosis capability to the methodology (Shi, Liu, and Lee, 2018; Gao et al., 2022; Wang et al., 2023) and remove redundant information (Sa et al., 2017a).

However, their feature evaluation process still mostly relies on thresholds that are determined either by the trend of previous operations' dissimilarity index or expert experience. The same quantitative transferability issue persists, and the recalibration of thresholds is, thus, required when the approach is to be applied to other turnouts.

2.2.1.3 Statistical feature-based approaches

When experts are not available and similarity analysis is deemed inefficient, statistical features can be analysed instead. First, an initial set of multiple features is calculated from data statistics. After that, only the relevant features will be retained for further evaluation, while the redundant ones will be eliminated (Eker et al., 2012).

Feature engineering is usually required to obtain the final feature subset. One of the most commonly used algorithms is Principal Component Analysis (PCA), which combines all initial features into Principal Components (PCs) that possess maximum variance, are not correlated to one another, and reside in a lower-dimensional and more abstract feature space (McHutchon, Staszewski, and Schmid, 2005).

Jin et al. (2015) and Shi, Liu, and Lee (2018) proposed two similar PCA-based methodologies for railway turnout FDD. They evaluated the Hotelling T^2 coefficient and Square Prediction Error (SPE), which were calculated from the resulting PCs, to detect faults for Alstom DK FBANE MET turnout systems. Once faults were detected, the former applied an unsupervised Self-Organising Map (SOM), and the latter applied a supervised K-nearest neighbours (KNN) algorithm for fault diagnosis. The results from both studies showed that the supervised classifier would provide superior accuracy.

Supervised feature engineering algorithms have been shown to facilitate better classification than unsupervised ones as well (Márquez, 2010; Guzman et al., 2018). The Linear Discriminant Analysis (LDA) algorithm is an example that linearly combines

all initial features by maximising the separation of the transformed features between classes (Ou, Xue, and Cui, 2019). Another algorithm called maximal-relevancy and minimal-redundancy (mRMR) was also proposed by Wang et al. (2023). As the name implies, the final feature subset would only contain features that had the greatest correlation with the class labels and the least redundancy among them. Both have been shown to facilitate superior FDD performance compared to the PCA algorithm.

Besides the evaluation of time-domain features, analysis of frequency contents extracted from high-frequency measurements, such as audio (Sun et al., 2020; 2021; Cao et al., 2021; Sun, Cao, and Li, 2022) and vibration (Cao et al., 2023), has gained more attention as well (Liu et al., 2022), where Lee et al. (2016) was the first article to introduce audio diagnosis to classify obstruction and slackened lock nut faults.

Applications of time-frequency domain features were also displayed by Asada (2013), who focused on the drive rod maladjustment at various severity levels. Here, the Discrete Wavelet Transform (DWT) was used to extract features, and then a multi-class Support Vector Machine (SVM) would classify them. The methodology achieved high accuracy when it was applied to both the Japanese NTS turnouts that use an induction motor as their main drive (Asada and Roberts, 2011) and the others that use a DC motor as well (Asada, Roberts, and Koseki, 2013; Asada and Roberts, 2013).

Chen et al. (2020; 2021) successfully reiterated this exact methodology, which was applied to ZD6 turnout systems commonly used in Chinese railways. The DWT procedure was identical to that of Asada's approach. However, the feature evaluation algorithm was changed from the SVM to a two-layer artificial neural network (ANN).

Although these supervised classifiers (such as SVM, ANN, Mixture Discriminant Analysis (Chamroukhi et al., 2008), Bayesian Network (Wang et al., 2017), and KNN)

and supervised feature engineering can achieve high FDD accuracy, they can only be applied when the available data are pre-labelled with multiple known health states (Li et al., 2021; Narges et al., 2022). Due to the lack of failure occurrences in the in-field turnout systems and the inability to record early fault developments, fault simulations are the only options that can provide these data in a deterministic way. As a result, the supervision level of the FDD systems would mostly depend on the comprehensiveness of these available training datasets and their corresponding health labels.

Because FDD systems should be able to acknowledge the early development of faults to facilitate timely maintenance planning, the indication of severity level is, therefore, also essential for practical use (Chen et al., 2021; Wang et al., 2023).

This requirement led to a series of research studies by Eker and his team that focused on slide chair fault developments by incrementally increasing the number of contaminated slide chairs (Eker et al., 2011). Here, they unsupervisedly clustered the statistical features to “mine” the optimal number of degradation levels (Camci et al., 2016; Atamuradov and Camci, 2017). However, using different clustering techniques or different criteria (e.g. Calinski-Harabasz and Silhouette) yielded different results, even when the same data were analysed (Atamuradov et al., 2018a; 2018b). Hence, the identification of fault severity was proven to be highly sensitive to the algorithm selection as well (Atamuradov et al., 2019; 2020).

Then, a similar approach was presented by Bian et al. (2018). Instead of one, they considered a total of six fault modes. The most representative features in this study were selected based on the correlation coefficient, Fisher discriminant criterion, and kernel PCA algorithm (Bian et al., 2019). The results emphasised the degradation differences of each fault mode that must be considered separately; see Figure 2-16.

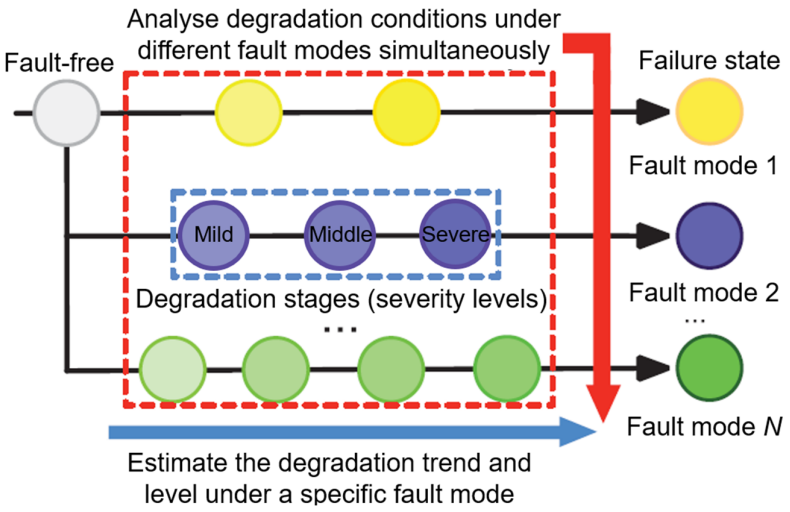


Figure 2-16 Different degradation trends of each fault mode (Bian et al., 2019)

Nevertheless, in practice, information on the corresponding fault mode may not be available in the maintenance database; only separation between data from failed and fault-free throws is possible (Böhm, 2017). Therefore, when data from in-field turnouts are evaluated, either health index analysis or anomaly detection (Vileiniskis et al., 2016; Sa et al., 2017b; Guo et al., 2020a) would be employed instead.

Ardakani et al. (2012) proposed a calculation of the health index based on the Hotelling T^2 coefficient of energy-related features. The results showed that the health index trend corresponds well with previous maintenance records. Soares et al. (2021) proposed another similar health index and an application of the unsupervised k-means clustering technique for HW2122 turnouts. The methodology could also effectively isolate failed throws by using the clusters of the calculated health index.

Frequency-domain features of motor current measurements extracted by the Fast Fourier Transform (FFT) algorithm were also employed by Mistry et al. (2019). However, instead of directly evaluating these frequency contents, they observed the performance of linear curve-fitting between the logarithm of frequency bin values and their FFT amplitudes. They concluded that the curve-fitting's residual standard error

(RSE) could indicate the fluctuation in motor current, which may reflect developing wear in the system. Thus, even if their data were completely unlabelled, degradation in each turnout could be observed via its RSE trend, and the optimal threshold could be set individually, as shown in Figure 2-17 (Mistry, Lane, and Allen, 2020).

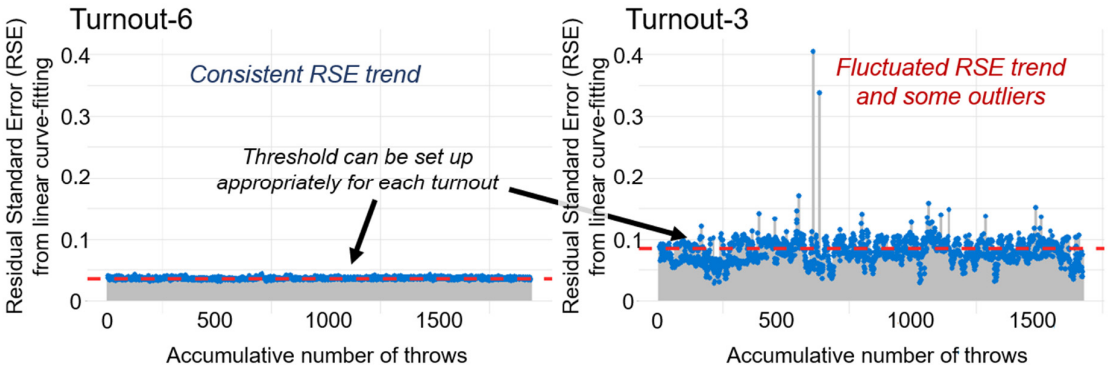


Figure 2-17 RSE trends from two different turnouts of the same type (Mistry et al., 2019)

One of the contributing factors that influences degradation characteristics, as shown in Figure 2-17, is the surrounding climates (Iwasawa et al., 2015; Jin et al., 2015; Letot et al., 2015; Wang et al., 2017). Guzman et al. (2019) were aware of this, so they compensated for the effects of differing temperatures first before calculating the anomaly score. The results showed that trends of the anomaly score correspond well only for turnouts that possess tight feature distributions (Guzman et al., 2020). As a result, they highlighted the importance of consistent system behaviours of the same health state that are critical for every FDD development by demonstrating examples of two features extracted from two turnouts that were fault-free, as shown in Figure 2-18.

From the plots of two features against the temperature bins, they concluded that a “good” dataset should possess well-defined features for each operational condition, like the black data, not as widespread as the red ones. Hence, a set of features that is optimal for one turnout may not be for the others, resulting in the transferability issue of these approaches in selecting the most representative features for each turnout.

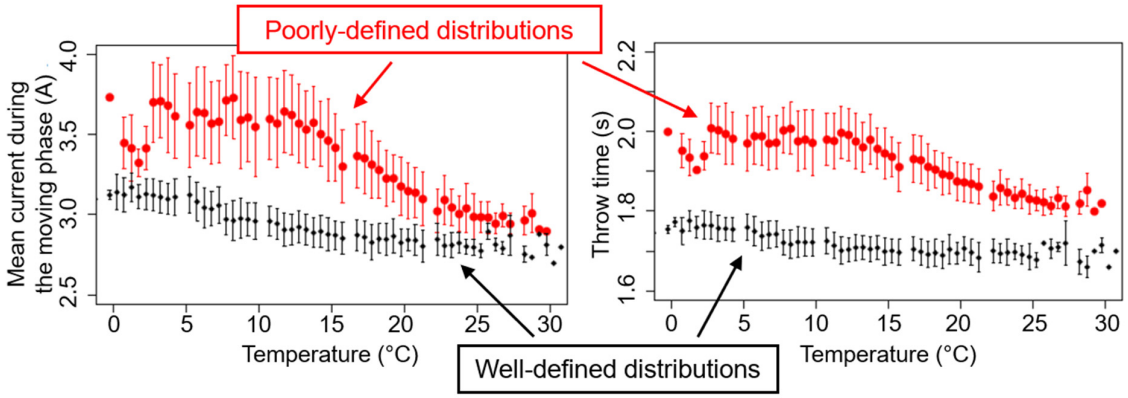


Figure 2-18 Distributions of "mean current during the moving phase" and "throw time" features extracted from two different turnout systems of the same design (differentiated by colours) when they were both in known fault-free conditions (Guzman et al., 2019)

All in all, the development of statistical feature-based FDD approaches requires the extracted features to be consistent with no to little overlapping distribution (Bian et al., 2019; Ou et al., 2020; Ladj, Bouamama, and Toguyeni, 2022; Hu et al., 2022). This causes the initial feature selection to affect FDD performance greatly. Moreover, there is no unified standardisation in developing this kind of FDD approach, but even slight changes in any aspect of the methodology may alter the diagnostic results significantly, although the same data are evaluated (McHutchon, Staszewski, and Schmid, 2005; Zhou et al., 2016; Atamuradov et al., 2020). Because there is also no "one-size-fits-all" approach, previous researchers mostly justified their methodology by comparing its performance to other methods (Wang et al., 2017; Chen et al., 2021).

The feature evaluation of statistical feature-based approaches normally applies machine learning techniques (Hu et al., 2022). One of the important requirements for their successful implementation is a balanced training dataset. However, that can only be obtained via fault simulations in laboratories. When data from in-field turnouts are analysed instead, data from faulty throws are always the minority and usually lack detailed labelling (Kim et al., 2016; Mistry, Lane, and Allen, 2020; Ou et al., 2020).

Several data augmentation techniques have been proposed to balance the dataset first before the evaluation (Zhou et al., 2016; Ou et al., 2020; Soares et al., 2021). However, the additional data were all artificially created based on the existing ones, which cannot cover the entire horizon of the turnout's characteristics.

Finally, although the initial features are usually intuitive and simple to compute, feature engineering and evaluation processes will adversely impact their interpretability (Hu et al., 2022). The former may transform the initial features into more abstract ones for better distinguishability. Meanwhile, all machine learning-based algorithms used for evaluation only attempted to map the final feature subset to the health state label in an abstract mathematical sense. Both reduce the explainability of the methodology greatly and usually lead to untraceable diagnosis results. Additionally, these sophisticated algorithms are usually not familiar to the technicians, making it difficult to justify them for safety-critical applications and causing a lack of active interactions.

Therefore, the trained FDD methodology that is best suited for one turnout may not provide the same level of performance for the others. It would require a new training dataset and a re-training process, resulting in the transferability issue.

2.2.1.4 Deep learning-based approaches

The previous approaches need a manual process to extract the initial features, which is one of the most critical steps that determines their FDD performance limit. Nevertheless, if the developers do not have any experience with the available dataset or when it becomes challenging to extract those representative features manually, deep learning-based approaches could be applied instead.

Deep learning algorithms implement deep-structured neural networks (NNs), which contain multiple and non-linear hidden layers. They can convert the original

measurements into abstract features that are best suited for the assigned objective function automatically (Zhang et al., 2018). An example of a Convolutional Neural Network (CNN), which is one of the deep learning models, is shown in Figure 2-19.

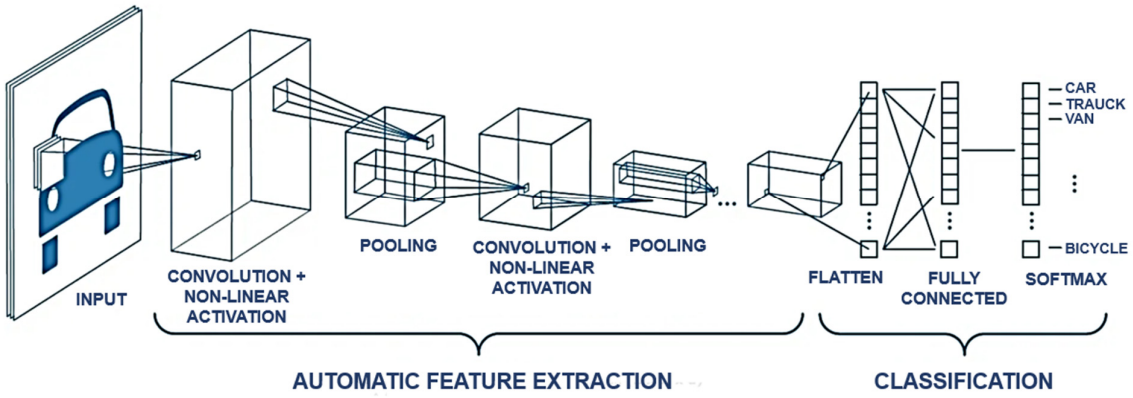


Figure 2-19 Illustration of CNN's deep architecture for image processing (Raghav, 2018)

Previous applications were used to evaluate unstructured "big" data containing latent abstract features, such as text (1-D), image (2-D) (Zhang et al., 2018), and video (3-D) (Ye et al., 2022). For 1-D applications, an autoencoder has been applied to analyse time series data. Essentially, it is an unsupervised feature reduction technique that takes every sample in the measurement as its initial features (Li et al., 2019). Instead of using linear transformation, autoencoders are non-linear, which allow them to capture latent and more complicated features better than PCA (Li et al., 2022).

Examples of complicated autoencoder applications have also been proposed as follows. The encoder and decoder of the autoencoder employed by Guo, Wan, and Ye (2020b) were structured from k-layer neural networks instead of one. A study by Guo et al. (2020a) analysed the reconstruction errors of the autoencoder instead of the extracted latent features. Here, they also included a few data obtained from faulty operations into the training optimisation. Guangwu and Rong (2021) proposed a Deep Belief Network (DBN) to enable an application of the three-layer deep autoencoder architecture, which further reduced the feature dimensionality for better evaluation.

As previously mentioned, the CNN is another successful and commonly used deep learning technique, but it requires a balanced dataset to enable effective fault diagnosis (Li and Fei, 2022). Shi, Du, and Yao (2023) proposed a deep CNN which employed several advanced optimisation algorithms to extract the optimal features. However, due to the scarcity of data from faulty operations, only fault detection was achieved when a Support Vector Data Description (SVDD) algorithm was applied.

All in all, deep learning-based approaches are analogous to feature engineering since they both attempt to classify data into different health states based on a set of abstract features. The major difference is that the deep learning-based approaches remove the manual feature extraction procedure, where initial features are extracted automatically by a deep data analysis architecture instead. However, this requires a lot more data to recognise complex patterns hidden inside them (Hu et al., 2022).

Due to such complete data reliance, overfitting to a specific dataset is the main drawback of deep learning-based approaches. Therefore, when two datasets obtained from the same health state are sufficiently different, the successfully trained framework may not perform desirably for both of them. Besides, the selection of its architecture structure and the associated hyperparameters can also impact the FDD performance.

Even with such complications, optimally trained deep learning-based methods usually outperform all others because the automatic feature extraction removes subjectivity from human intervention and chances that informative but latent features are not being extracted. However, they hold the least interpretability and traceability due to their sophisticated data analysis architecture. Due to this lack of explainability, deep learning-based approaches are the most difficult to justify and practically implement for safety-critical systems (Jimenez et al., 2020; Hu et al., 2022).

2.2.1.5 Conclusion of data-driven methodologies

Data-driven FDD approaches evaluate features that are directly extracted from the available measurements without consideration of the relationships between the system's inputs and outputs. Since the empirical knowledge of system behaviours can only be attained through these data, their comprehensiveness and quality will dictate the performance limit (Shi, Liu, and Lee, 2018; Chen et al., 2021; Li et al., 2021).

The main concern of these FDD approaches is transferability, as there is no integration of other information apart from the data (Bukhsh et al., 2019). That is, a methodology optimally trained from one turnout may not perform well for others. The chances are even more likely if their environmental and working conditions differ significantly, resulting in high dissimilarities in their measurements (Du et al., 2019).

Because every practical FDD system should also be able to acknowledge the early development of faults, this thesis will develop an FDD methodology based on a comprehensive set of data obtained from fault simulations while also proposing the integration of a transferability mechanism to alleviate the transferability issue.

Interpretability and traceability also influence the practicality of the developed FDD methodology. These depend on how abstract and complicated it is. A simple approach that provides accurate results tends to be more justifiable for safety-critical applications (Bukhsh et al., 2019; Gao et al., 2022). Such an FDD methodology would allow the users to re-evaluate the result by tracing it back to the original measurements. This will greatly facilitate an active interaction between the users and the FDD systems.

According to these challenges, the implementation of data-driven philosophies into the FDD developments for railway turnouts is summarised in Table 2-2.

Table 2-2 Summary of comparison analysis on the previously developed data-driven FDD methodologies for railway turnout systems

Data-driven methodologies	***Pros***	***Cons***
Expert knowledge	Features are simple to calculate and will not get transformed. So, they preserve all interpretations. Since they are normally considered during maintenance, the approaches are easy to follow, making them traceable.	A deep understanding of the monitored systems is mandatory, whereas experts have become less available nowadays. Features and optimal thresholds used are also particularly practicable only for the turnouts that the experts have experience with.
Similarity analysis	These approaches can provide visual interpretation to the users based on waveform comparison. Although they require less expert knowledge, these approaches are still simple to understand and can be thoroughly traced.	Dissimilarity indices are normally evaluated by thresholds, but they are always direction- and turnout-specific, requiring individual setup for each circumstance. Also, signal noise can adversely influence its performance.
Statistical features	The computational overhead is reduced due to the feature engineering. Development of these approaches requires even less expert knowledge of the monitored system but more on the data analysis side. Several techniques are already available to apply. With proper optimisation, they can achieve high accuracy.	They require the initial features to be extracted manually, which can lead to a risk of information loss and redundant features. The machine learning-based evaluation is difficult for technicians to understand. Also, feature engineering can render the interpretability of the final feature subset difficult.
Deep learning	Expert knowledge is no longer required during the development. It is most suitable if the available data are unstructured. The most effective features of the dataset are guaranteed to be extracted	These approaches require a tremendous amount of pre-labelled training data to recognise the patterns residing in them. So, they are usually prone to be overfitted to the training dataset.

Data-driven methodologies	***Pros***	***Cons***
Deep learning (Cont.)	automatically through intensive learning algorithms, even though they are latent inside the dataset.	Also, they are not interpretable and traceable, as every procedure is conducted in a “deep” data analysis architecture throughout the process.

Based on this summary, it is concluded that employing solely any one of these principles for FDD developments is not practical for railway turnout systems. The investigation is now continuing to model-based methodologies.

2.2.2 Model-based methodologies

The term “model” translates to a mathematical model that aims to capture the relationships between a system’s inputs and outputs (Pecht and Jaai, 2010). Once a representation of the monitored system is formulated, the available features that can be extracted from it include (1) the differences between the measured outputs from the system and the model estimations, called “residuals”, and (2) the characteristics of the derived models. Residuals produced from the validated model should be negligible when the system is fault-free. In contrast, they should be substantial enough to be informative FDD features when the system becomes faulty (Chen and Roberts, 2006).

Unlike data-driven methods that usually examine one specific measurement, model-based approaches always consider more than one (i.e. at least one input and one output). This characteristic facilitates the isolation of effects from measurement noise during residual calculation, which can be filtered for better detection of faults in the system or used to detect sensor deteriorations (Zattoni, 2006; Ladj, Bouamama, and Toguyeni, 2022). Nevertheless, a training dataset is still required as its extracted features will determine the decision boundaries in the feature space.

The development of mathematical models can be varied based on how much the system knowledge is integrated. In this thesis, model-based methods are divided into two sub-categories, namely black box and first principle model-based approaches.

2.2.2.1 Black box model-based approaches

This kind of model takes the data fitting quality of its estimation to the highest priority, such that the model's objective is to estimate the system's outputs accurately when the inputs are known. Any combination of arbitrary mathematical functions can be chosen to formulate the model structure (e.g. linear, polynomial, exponential, and trigonometric). With such flexibility, the model can capture the system's complex and non-linear relationships accurately using abstract mathematics (Roberts et al., 2002).

Since the model structure does not integrate a priori system knowledge, these models will not possess any physical interpretations. When their model structure has been chosen, its parameters will be obtained through an optimisation process based on the system measurements. Without any other system-related information included during the model development, residuals are the only informative FDD features.

Only a few studies have applied this concept to turnout FDD. First, Roberts et al. (2002) employed third-order polynomial functions to model the relationship between the air pressure, the switch rail's velocity, and the total air volume collected from four-foot electro-pneumatic turnouts. They divided each throw into six phases and assigned polynomial models individually. This segmentation enhanced the fault sensitivity while not overly complicating the computational process. The residual evaluation with both the preset thresholds and an adaptive neuro-fuzzy inference system (ANFIS) allowed the methodology to detect and isolate multiple fault modes.

The qualitative aspects of this methodology (i.e. the polynomial model structure and the rule base in ANFIS) could be transferred to other similar electro-pneumatic turnouts. However, the quantitative residual generation and the preset thresholds had to be individually re-calibrated first. This resulted in a transferability issue similar to the findings of Silmon and Roberts (2010), who proposed the QTA algorithm.

Similar FDD research was proposed by Bai (2010), who attempted to develop a generic methodology for railway single-throw mechanical equipment (STME) based on multiple abstract mathematical models; see Figure 2-20.

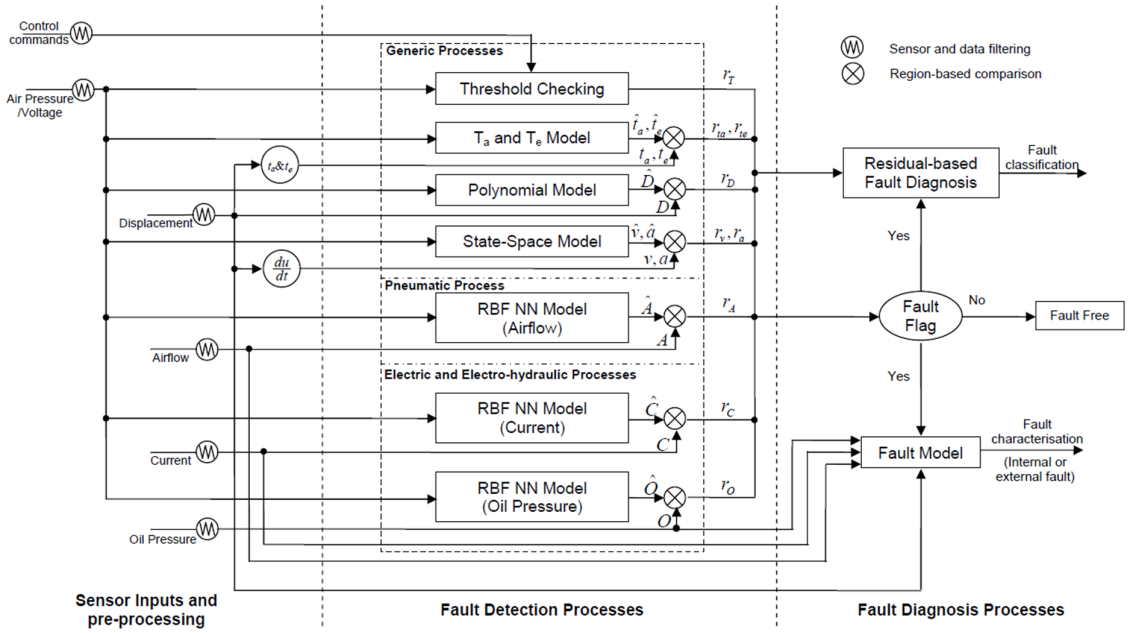


Figure 2-20 Generic FDD methodology for several railway STME assets (Bai, 2010)

The process started with an adaptive threshold checking of the input signal when the control command was sent. The activation delay and throw times calculated from the displacement were then compared to those from an exponential estimation of the system's input. Finally, the polynomial and state-space models were developed to replicate the system's dynamics. The former focused on the displacement, and the latter on the velocity and acceleration. Several radial basis function neural network

(RBFNN) models were also developed to classify whether the detected fault was internally or externally caused (Bai, Roberts, and Goodman, 2008).

To the author's knowledge, Fidali, Wojciechowski, and Pelka (2018) presented the latest black box model-based FDD development in this research field. Here, a simple autoregressive with exogenous input (ARX) and ANN was used to represent the relationship between the motor current and the output torque of an EBI Switch 2000 turnout. With a set of arbitrary thresholds derived from previous fault-free throws, the residual evaluation of both models was capable of sufficiently accurate fault detection.

In conclusion, black box models aim to replicate the relationships between a system's inputs and outputs by using abstract mathematics. This property allows them to mimic complicated system dynamics, such as the railway STME (Bai, Roberts, and Goodman, 2008) and jointless track circuits (Zhao et al., 2013) accurately. However, since the system knowledge is not integrated into the model structure, it does not contain any physical interpretation. In addition, even if the same model structure is applied to different turnouts of the same type, a similar quantitative transferability issue will still exist, requiring individual re-training for each model (Roberts et al., 2002).

2.2.2.2 First principle model-based approaches

Analogously, first–principle models also attempt to replicate the relationships between a system's inputs and outputs. However, they **explicitly** integrate system knowledge into the model development as well, especially its governing physics (Jimenez et al., 2020). For linear time-invariant (LTI) systems, their model structure is in either differential or difference equations, which can be rearranged into a state-space or a transfer function model (Ding, 2008; INNOTRACK, 2009b). Residuals from validated models should be the lowest when the system is in fault-free states, robust to uncertainties (e.g. signal noises and modelling errors), and sensitive to faults.

To the author's knowledge, only Zattoni (2006) has applied this concept to a physical working turnout system. First, a linear state-space model was developed from equations that governed the behaviours of a P80 turnout. Since the system behaved linearly during the moving phase, this model could accurately replicate the dynamics of that interval only. Figure 2-21 validates this assumption by comparing the motor current from the P80 turnout (the solid line) to the model estimation (the dashed line).

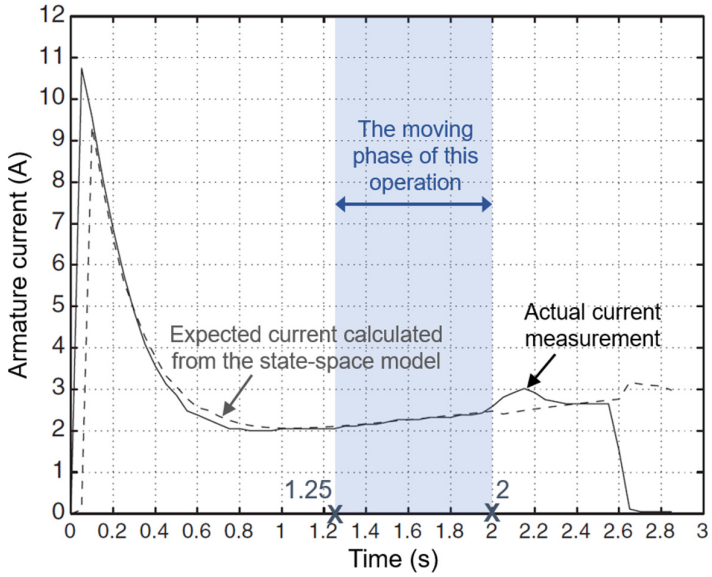


Figure 2-21 Comparison between the measured current (the solid line) and the estimation calculated from a first principle model (the dashed line) of a P80 system (Zattoni, 2006)

From the figure, the moving phase was identified to start from 1.25 to 2 seconds. With a sampling time of 0.05 seconds, this phase lasted for 16 samples. This number was used to design finite impulse response (FIR) models to calculate residuals based on the H_2-norm criterion. Their results demonstrated that the residuals would only be sensitive to additional terms related to faults, not other disturbances, achieving the desired characteristic of "good" residuals. After that, their RMS values were evaluated, allowing the classification of simulated bearing faults at multiple severity levels.

Because first principle models can be designed to capture system dynamics entirely and accurately, most of the research studies employed this philosophy to

facilitate controller and fault-tolerant designs instead. For example, Dutta et al. (2018; 2020a) developed a simulation model to design a self-adjusting controller for an HPSS system, making it tolerant to slight misalignment faults. Then, Li et al. (2022) simplified their complicated first principle model of the switch rail's dynamics to develop a fault-tolerant controller for a high-redundancy actuator. Meanwhile, Olaby et al. (2021) used them to create a closed-loop controller for a REPOINT-Light turnout system.

The most recent research implementing first principle models for turnout FDD was proposed by Hamadache et al. (2019b). First, they developed a simulation model of an electro-mechanical turnout in Simulink®, and friction faults were then simulated by incrementally increasing "the total system's viscous friction coefficient" parameter. Their results demonstrated that the accumulative residuals were more distinguished than the raw unprocessed residuals, thereby facilitating better fault detection.

Compared to all other methodologies, first principle model-based methods are the least common for turnout FDD, which corresponds with the finding of Hamadache et al. (2019a). The primary reasons are as follows.

First, railway turnout operations are not strictly linear due to the unlocking and locking phases. Second, the switch rail's bending dynamics during the moving phase contribute to time-varying loads. Since first principle models should perfectly duplicate fault-free turnouts, their model structure has to be non-linear and overly complicated as well, whereas obtaining its parameters is also extremely challenging.

Matei et al. (2015) was the only study that considered both non-linearities in the turnout model development. Nevertheless, similar to the models developed by Zattoni (2006) and Oyebande and Renfrew (2002; 2007), they were not entirely validated because their model output estimation matched that of the system in some phases

only. Furthermore, the system parameters provided by the manufacturers may be incomplete and no longer correct. If they were to be obtained through intensive field experiments, they would not be applicable to all distributed in-field turnouts, resulting in transferability and applicability issues (Hamadache et al., 2019a).

Ideally, suppose an accurate first principle model for a specific turnout system can be developed. In that case, it can be regarded as an “analytical redundancy”, which can be used to study the system’s behaviours in any health state in-depth instead of actually conducting experiments on it (Pecht and Jaai, 2010; Dowdeswell, Sinha, and MacDonell, 2020). However, this perfect scenario has yet to be achieved for railway turnout applications, so this potential benefit has not been investigated.

2.2.2.3 Conclusion of model-based methodologies

Model-based FDD approaches extract features from mathematical models that try to replicate the relationships between a system’s inputs and outputs (Márquez, Roberts, and Tobias, 2010; Hamadache et al., 2019a; Hu et al., 2022; 2023). Such philosophies have not received the same level of attention compared to data-driven counterparts. This is due to the effort required for model development (i.e. determining the optimal combination of model structure and its parameters) and considerations of multiple system measurements (i.e. at least two are necessary for SISO systems).

The model structure of black box models can be chosen arbitrarily. Then, the model parameters are determined based on the minimisation of the estimation error (Ljung, 1999). So, they can be designed to be capable of mimicking complex and non-linear system dynamics (Roberts et al., 2002). Due to the complete data reliance, even if the model structure is fixed, using training data from different turnouts would result in different optimal parameters. This characteristic results in a similar transferability issue in which black box models have to be adjusted individually first. Lastly, because

they are mathematically abstract, these black box models will not retain any physical interpretation of the original system (Fidali, Wojciechowski, and Pelka, 2018).

In contrast, the model structure and parameters of first principle models, when the system is in fault-free conditions, must be known prior to the residual generation. This philosophy assumes that “deep” system knowledge is available. Therefore, it is often regarded as a “white box” where all related system dynamic characteristics are retained in the model. However, creating an accurate mathematical model of complex systems can be challenging. Therefore, they have not been extensively studied in this research field. According to these characteristics, the pros and cons of applying these model-based philosophies for railway turnout FDD can be summarised in Table 2-3.

Table 2-3 Summary of comparison analysis on the previously developed model-based FDD methodologies for railway turnout systems

Model-based methodologies	***Pros***	***Cons***
Black box model	Model structure can be arbitrarily chosen from any mathematical functions that are simpler than the actual system’s equations describing its complicated dynamics. Identification of the optimal model parameters can be done by minimising the model estimation error, where any optimisation strategy can be implemented.	Black box models bear no physical interpretation of the system. Therefore, an optimal model structure for one turnout may not be applicable to others. Data-driven approaches provide similar limitations and FDD performance. However, they can be conducted with one system variable, whereas black box models require at least two.
First principle model	The detailed underlying physics of the system is integrated into the model, ensuring the transferability of the model structure from one turnout to another. Simulation of every health state can be	The model structure can be overly complicated with a substantial number of parameters. As a result, not all optimal parameter values may be available, leading to models not being able to be

Model-based methodologies	***Pros***	***Cons***
First principle model (Cont.)	conducted on the validated model instead, as it is an “analytical redundancy” of the system.	fully validated. Even if they were, they are still not applicable to all turnouts that are widely distributed along the railway.

Based on the reviews of existing methodologies, every principle has its unique limitations that prevent it from being a practical solution for railway turnout FDD. Therefore, the following section will highlight these practical issues and finally propose an alternative FDD principle that aims to solve these problems simultaneously.

2.2.3 Summary of limitations of existing FDD methodologies

Sophisticated FDD methodologies were first proposed to replace thresholding-based approaches, where raw measurements are directly compared to the preset thresholds. This was due to their frequent inaccurate results that negatively affect the turnout’s reliability. As a consequence, the first requirement that every FDD system must possess is **sufficiently high diagnosis accuracy**.

When they are applied to railway turnouts to prevent failure occurrences, these FDD systems must be capable of acknowledging early fault developments. Also, due to complicated turnout structures, they can be deteriorated in many ways. As a result, fault detection alone is not adequate to facilitate effective maintenance planning. These FDD systems should be able to **classify multiple fault modes at different severity levels** to produce more detailed diagnostic information (Goodall and Roberts, 2006).

The above capabilities can only be achieved if the training dataset is thoroughly labelled and comprehensive. Such data are exclusively available from fault simulations because data collected from in-field turnouts are usually imbalanced and only contain failure scenarios. Furthermore, due to the distributed nature of railway turnouts that

are operated in different environmental and working conditions, an FDD methodology optimally designed for one of them may not be entirely transferable to the others. As a result, another requirement for practical FDD systems is **a transferability mechanism** that only requires the least re-training process and still enables them to be sufficiently optimal for all other instances of similar turnouts.

Regardless of their performance, FDD systems can only advise the technicians by providing diagnosis information to support the maintenance planning process. They still manually decide when and how the maintenance should be conducted on a particular turnout. Therefore, using abstract mathematics or complicated procedures in the FDD algorithms would only allow them to “blindly” engage the results, as the rest of the algorithm would be too abstract and beyond their comprehension.

Since this thesis focuses on FDD applications for safety-critical railway turnout systems, the implemented algorithms should be **traceable and interpretable** to the users. These characteristics would allow such advisory systems to be justifiable to the infrastructure managers. At the same time, they will also improve confidence and trust in the FDD results, as they can be re-evaluated by the users if required.

Most of the previous FDD developments for railway turnouts only aimed to solve the original accuracy problem. However, when all other requirements were in place, all of them failed to be considered practical solutions for this particular application. To support this statement, Table 2-4 elaborates on the evaluations of all previous FDD philosophies based on the three proposed requirements.

Table 2-4 Evaluations of each FDD category of the three proposed requirements

FDD category	**Requirements for practical railway turnout FDD systems**		
	Detection and diagnosis	*Transferability*	*Traceability and interpretability*
Expert knowledge	It depends on the level of expert knowledge available during the FDD development.	The recommended features are only effective for particular turnouts they have experience with.	Normally, the proposed features have already been approved and used during the maintenance process.
Similarity analysis	Normally, similarity analysis conducted on the entire operation can only detect faults. However, employing segmentation could facilitate fault diagnosis capability as well.	Fault-free reference data would be specific for one turnout only. Inevitable differences between turnouts prevent it from being directly transferrable.	Comparing two graphs quantitatively or their trend changes is similar to how humans would analyse them. Hence, the process is intuitive and also provides visual interpretability.
Statistical feature	Features extracted from segmented data are more capable of diagnosing faults compared to those derived from the entire operation.	Patterns of features extracted from fault-free and faulty throws can be different from one turnout to another. Data collection for re-training is required.	Feature engineering processes usually transform initially intuitive features into more abstract ones, potentially removing their interpretation.
Deep learning	Apart from the required comprehensive training dataset, including prior segmentation before the feature extraction can facilitate diagnosis.	The features are completely latent and highly specific to the training dataset. Hence, the deep learning model is highly likely to be overfitted.	Both feature extraction and evaluation reside in a “deep” data analysis architecture. Such processes are mostly not familiar to the technicians.
Black box model	It depends on the consistency of changes in residuals and the	The model parameters are uniquely obtained for each turnout from	Since they employ abstract mathematical functions, these

FDD category	Requirements for practical railway turnout FDD systems		
	Detection and diagnosis	*Transferability*	*Traceability and interpretability*
Black box model (Cont.)	evaluation algorithm used. Multiple residuals from multiple variables are also better at facilitating diagnosis.	their data. So, they are not transferrable. Since the model structure is abstract, it may not be transferrable as well.	residuals are abstract and do not hold any physical interpretations.
First principle model	Sufficiently detailed residual evaluation and consistent changes in residuals from each health state can facilitate the diagnosis capability.	The model can be greatly complicated, but its structure is transferrable to every similar instance, but model parameters are usually not.	Since the originality of residuals is known from the first–principle models, the users can intuitively relate certain residuals to specific fault modes.

According to Table 2-6, it can be concluded that *no previous FDD developments attempted to consider these requirements simultaneously*. As a consequence, the FDD methodology to be developed in this thesis is going to consider all three requirements. Parameter estimation-based FDD methodologies, which have not been reviewed yet but have shown lots of potential in many simulation studies, will be proposed, and how they hypothetically meet these requirements and overcome the challenges will be elaborated in the following section.

2.2.4 Proposed feature extraction technique: Parameter estimation

Parameter estimation approaches consider another available feature that can be extracted from the system's mathematical representation, which is the model's inherent characteristics (Ljung, 1999; Overschee and De Moor, 2012).

This concept combines both first-principle and black-box model philosophies. That is, the model structure is known a priori from the system's underlying physics,

whereas its parameters are optimally calculated based on the system measurements (Dixon, 2000; Dutta et al., 2020b). It utilises the qualitative transferability of the model structure and allows the flexibility of the model parameters to be individually adjusted.

The two previous model-based methods require a model that must accurately estimate the system outputs to produce reliable residuals (Patcharaprakiti et al., 2010). In contrast, parameter estimation-based methods focus on the changes in the model parameters when the system becomes faulty instead. Suppose the parameter changes corresponding to each health state are consistent. In that case, they will facilitate fault diagnosis capability in the approach. Therefore, being able to detect these parameter changes is the primary objective, not achieving high accuracy in the output estimation (Dutta et al., 2022; Ugwiri et al., 2022). Since such features can be obtained from a simpler model that adequately represents system dynamics, some complexity and non-linear aspects of the system can be neglected or simplified (Zhang et al., 2014).

Apart from the qualitative properties of the model structure, the relationships between the qualitative changes of model parameters and each health state are also transferrable among a group of similar turnouts (Dixon and Pike, 2002; Antong, Dixon, and Ward, 2017). Such information can be obtained and serve as “empirical” expert knowledge. This property is intuitive as the model structure is derived from the physical characteristics of the system, so its parameters would hold some degree of physical interpretation (Isermann and Balle, 1997; S-CODE, 2019). So, thorough interpretability and traceability of the parameter estimation-based FDD approaches would be dictated by the evaluation algorithms that should be interpretable and traceable as well.

In conclusion, parameter estimation-based FDD methods meet the proposed requirements for railway turnout applications as follows:

- The approaches have an inherent transferability mechanism because the same model structure can describe the dynamics of other similar turnouts. Meanwhile, inevitable differences between them can be accounted for by individually re-estimating the model parameters based on each fault-free training dataset.
- The validation of the model structure can be done in a laboratory, where fault simulations can be comprehensively conducted on the test turnout. Therefore, the features that are capable of early fault detection and classification of multiple fault modes can be determined. Then, the resulting “empirical” knowledge base of relationships between the parameter changes and health states will be ideally transferrable to all instances of similar turnouts as well.
- Since the model structure is derived from the system’s underlying physics, its parameter estimates also hold physical interpretation. Therefore, this process ensures the interpretability of the approach’s feature extraction process. The level of complexity and abstractness of the following evaluation procedure will determine whether the interpretability propagates throughout the methodology.

Parameter estimation-based FDD approaches have been successfully applied in the past, including ones for electric machines (Wijewardhana et al., 2020; Dilys and Stankevič, 2022), a single-phase inverter (Patcharaprakiti et al., 2010), a residential HVAC system (Turner, Staino, and Basu, 2017), electro-mechanical (Janot, Young, and Gautier, 2017; Kim and Tahk, 2018) and electro-magnetic actuators (Forrai, 2017), lithium-ion batteries (Jiang et al., 2017), and centrifugal pumps (Ugwiri et al., 2022).

There have also been a few railway applications, including mechanical systems such as bogies (Alfi et al., 2011; Zhang et al., 2014) and their suspension components

(Zoljic-Beglerovic et al., 2018). Both of them first derived linear state-space models to simplify their system dynamics, which were highly non-linear, strongly coupled, and time-varying. The former proposed an extended Kalman filter (EKF) to estimate both the parameters and non-measurable states required for the estimation. Meanwhile, the latter proposed a sliding mode-based algorithm that could account for measurement noises and time-varying parameters. Due to their complicated natures, the practicality of both FDD approaches could not be validated with their respective physical system.

An application example for railway electrical systems was proposed by Yang, Zhang, and Gu (2019). First, they developed a cascade two-port network model of the ZPW-2000 jointless track circuits, which require functional compensation capacitors to tune their frequency response. Deterioration of these capacitors could be detected by investigating their capacitance and internal resistance. Since the approach required voltage and current measurements on the track collected from an inspection train, which was not available yet, it was only developed and validated through simulations.

An application for an electro-mechanical system was proposed by Dassanayake et al. (2009). A physical electric overhead train door was studied in this research. First, they developed two transfer function models, each representing the mechanical and electrical responses of the system, to be used as the model structures. To remove the noise resulting from signal derivative approximations, they applied linear integral filters (LIFs) and used recursive least squares (RLS) as their main estimation algorithm.

The results indicated that when faults were simulated on the test train door system, changes in the expected model parameters were noticeable. However, after the methodology had been successfully developed, it was not deployed to other similar train door systems. Although they did not validate the transferability, this work has shown much potential for practical use of the parameter estimation algorithms.

Parameter estimation has already been proposed to detect and diagnose faults in railway turnout systems, as presented by Dutta et al. (2020b; 2022). However, both developed methodologies based on data collected from complicated first principle simulation models of the turnouts instead of the physical ones. Faults were simulated by changing the model parameters, and the approach could successfully detect them.

Thus, this thesis will develop a parameter estimation-based FDD methodology based on data collected from a test turnout system in a laboratory. Data from both directions (i.e. normal-to-reverse: N-to-R and reverse-to-normal: R-to-N) in fault-free states and various fault modes will later be obtained and evaluated. A flowchart of the entire development of the proposed FDD methodology is shown in Figure 2-22.

2.2.5 Conclusions

The second half of this chapter evaluated the characteristics and limitations of previous FDD developments for safety-critical railway turnout systems. This literature review concluded that most techniques did not effectively meet all requirements for their practical applications. Only parameter estimation-based methodologies showed potential from previous simulation studies. This is because the parameter changes associated to different health states facilitate fault diagnosis of multiple fault modes and early fault detection. The parameter estimates also attain physical interpretation due to the use of system's governing physics in developing the model structures. The implementation of this technique will begin in the following chapter.

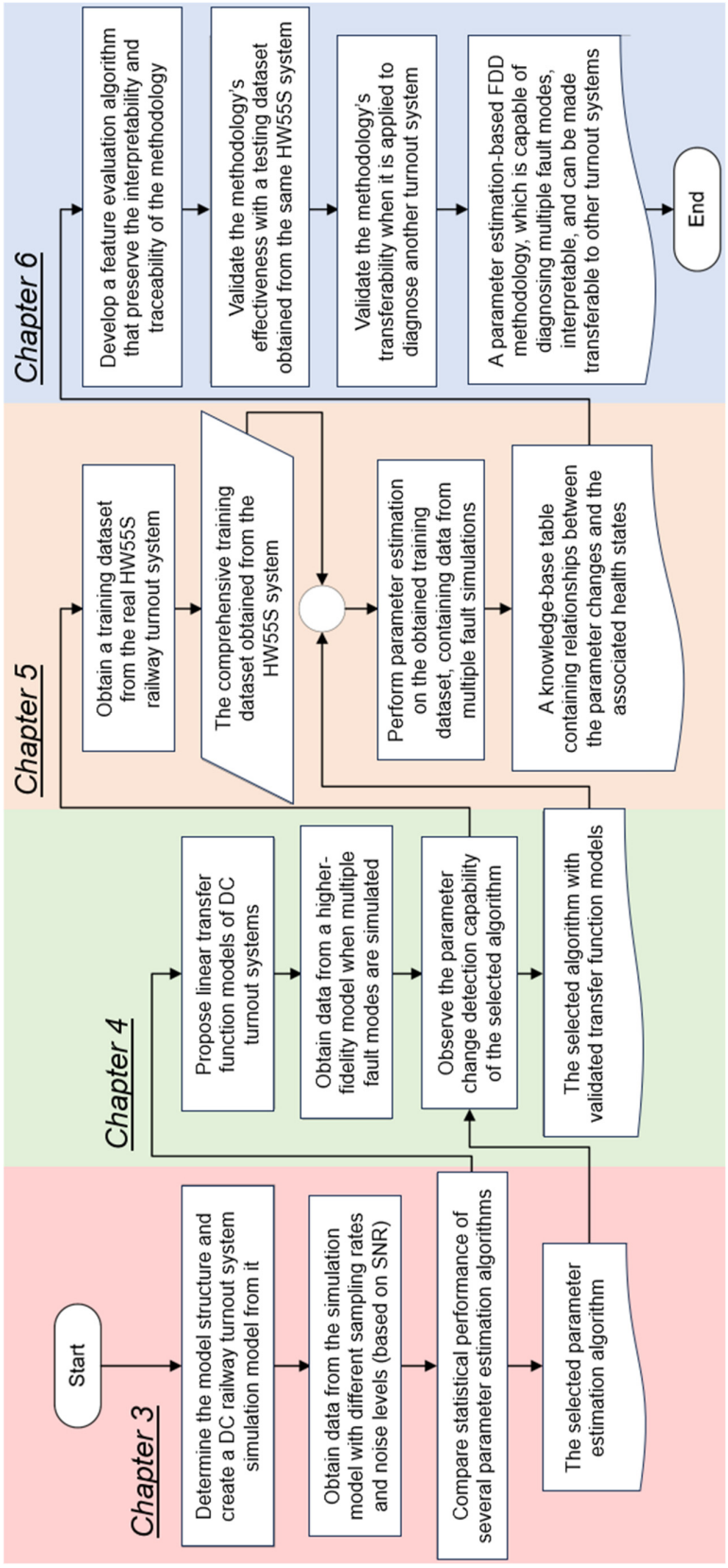


Figure 2-22 Flowchart illustrating the entire development process of the proposed parameter estimation-based FDD approach for electro-mechanical turnout systems

CHAPTER 3 APPLICATION OF PARAMETER ESTIMATION FOR RAILWAY TURNOUT FDD

Development of a parameter estimation-based FDD methodology for railway electro-mechanical turnouts begins in this chapter. First, the application of parameter estimation for FDD purposes will be described. Then, two categories of algorithms, namely indirect and direct, will be compared to facilitate the algorithm selection. At the end, a simulation model of electro-mechanical turnouts will be developed to evaluate the performance of the selected algorithm. The criteria include statistical properties of the parameter estimates, robustness to noise, and sensitivity to the sampling rate.

3.1 Parameter estimation for FDD purposes

Parameter estimation is an integral part of system identification, which attempts to replicate relationships between a system's inputs and outputs mathematically. The system identification framework will first select the optimal model structure and then estimate its parameters to minimise the discrepancy between the outputs of the system and the model estimation (Young, 1974; Frank, 1990). Consider the physical system, which can be mathematically expressed as follows:

System equations: $$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}\left(\mathbf{u}(t), \mathbf{x}(t), \mathbf{v}(t), \boldsymbol{\rho}(t), t\right) \quad \text{(3-1a)}$$

Observation (or measurement) equations:

$$\mathbf{y}(t) = \mathbf{g}\left(\mathbf{u}(t), \mathbf{x}(t), \mathbf{w}(t), t\right) \quad \text{(3-1b)}$$

where $\mathbf{x}(t)$ is the internal states vector of the targeted system;

$\mathbf{u}(t)$ is the external input vector that excites the system;

$\mathbf{y}(t)$ is the output vector measured from the system. Note that only $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are accessible for measurements;

$\mathbf{v}(t)$ and $\mathbf{w}(t)$ are the disturbance vectors affecting the states and measured outputs, respectively. They determine the system's stochastic properties, which can be described by their probability density functions (PDFs);

$\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are functions that provide structures of the governing equations describing system dynamics and measurement process, respectively;

$\boldsymbol{\rho}(t)$ is the parameter vector that characterises the system dynamics once the model structure of $\mathbf{f}(\cdot)$ is determined.

For FDD purposes, parameter estimation can be applied in two ways.

- First, if the model structure representing the system, as in Eq. (3-1), is known, the model will be able to represent the system accurately (Chow and Willsky, 1984; Söderström and Stoica, 1985). This property allows any health state to be simulated on the model instead of on the actual system. Then, a comparison between the outputs from faulty conditions and those from fault-free scenarios would produce **residuals** as effective FDD features.
- In contrast, if the exact model structure is unknown or too complicated to be implemented, their simplified representation can be applied instead. Since the model cannot represent the system accurately, a test system is still required. Fault simulations will then be conducted on it, where each health state should result in distinct values of **model parameters**. Suppose their changes are consistent. They would be considered reliable FDD features (Dixon, 2000).

Because the dynamics of railway turnouts are complicated (due to their locking mechanism and the switch rail's bending behaviours), an accurate model of them will be difficult to obtain. Hence, the FDD methodology to be proposed in this thesis will be one that *examines model parameter changes, which are related to fault developments in the system*. The following section provides an overview of parameter estimation

algorithms that have been used for FDD purposes in this manner. Then, a comparison is conducted to facilitate the algorithm selection that is adopted in this thesis.

3.2 Overview of parameter estimation algorithms

Theoretically, a set of differential equations is able to describe the dynamics of most physical systems. These continuous-time domain (CT) expressions are familiar to scientists and engineers, as they are physically interpretable (Garnier, Mensler, and Richard, 2003). Nevertheless, due to the advancement of digital signal processing, the trend of parameter estimation has been shifted to one that applied to their equivalent discrete-time (DT) expressions instead (Young, 1974; Ljung, 1999; Ding, 2008).

The following subsections outline two parameter estimation principles based on the type of model structure they are applied to (whether the original CT model or its DT equivalent). Each has unique properties and requirements, which are evaluated to facilitate the selection of the algorithm to be implemented for railway turnout FDD.

3.2.1 Indirect parameter estimation principle

As the name implies, the estimation is not directly conducted on the original CT model. It is conducted on DT equivalent instead (Jakeman, Steele, and Young, 1980). For this purpose, linear time-invariant (LTI) models are commonly used to set up the model structure, as they can represent linear dynamical characteristics while not being too complicated (Söderström and Stoica, 2002; Patcharaprakiti et al., 2010).

Referring to Eq. (3-1), assume that the dynamics of the targeted single-input, single-output (SISO) system can be described by a CT LTI model as follows:

System equation: $x(t) = \frac{B(s)}{A(s)} u(t)$ (3-2a)

Observation (or measurement) equation: $y(t) = x(t) + n(t)$ (3-2b)

where $x(t)$ is the theoretical "noise-free" output, which is solely influenced by the system input and not accessible for measurement;

$u(t)$ is the input that excites the system and is available for measurement;

$y(t)$ is the measured output which is contaminated by signal noises, $n(t)$;

$B(s)$ and $A(s)$ are numerator and denominator polynomials of the system's CT transfer function in the *s*-domain, respectively.

The transfer function describing the system dynamics, as in Eq. (3-2a), must be stable (i.e. no poles possess a positive real value) and either proper or strictly proper (i.e. the degree of denominator polynomial must be equal to or higher than the degree of numerator polynomial respectively). To prevent further simplification of this transfer function, the polynomials $B(s)$ and $A(s)$ should also be co-prime to each other (i.e. no common factors between them). Both polynomials can be expanded as follows:

$$A(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \ldots + a_{n-1} s + a_n \quad \text{(3-3a)}$$

$$B(s) = b_0 s^m + b_1 s^{m-1} + b_2 s^{m-2} + \ldots + b_{m-1} s + b_m, \text{ where } n \geq m \quad \text{(3-3b)}$$

From Eq. (3-2), *the CT model in the s-domain must be transformed into its DT equivalent first for indirect parameter estimation*. The resulting DT LTI model is:

$$y(t_k) = \frac{\beta(z^{-1})}{\alpha(z^{-1})} u(t_k) + e(t_k) \quad \text{(3-4)}$$

where $u(t_k)$ and $y(t_k)$ are the sampled system's input and output. The subscript "*k*" indicates the instant of that sample, where $1 \leq k \leq N; k \in \mathbb{Z}^+$, and *N* is the total number of data samples being evaluated;

$\beta(z^{-1})$ and $\alpha(z^{-1})$ are the "true" numerator and denominator polynomials of the DT transfer function representing the system dynamics. They are the conversion products from the CT *s*-domain to the DT *z*-domain.

$e(t_k)$ is the sampled error between the theoretical "error-free" output and the measurable one, i.e. $y(t_k)$. It includes every aspect that contributes to the mismatch, including modelling errors, measurement noises $n(t_k)$, algorithm performance, and the selected domain conversion process.

After the domain conversion, $\beta\left(z^{-1}\right)$ and $\alpha\left(z^{-1}\right)$ can be expressed as:

$$\alpha\left(z^{-1}\right) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots + \alpha_{p-1} z^{-(p-1)} + \alpha_p z^{-p} \tag{3-5a}$$

$$\beta\left(z^{-1}\right) = \beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \ldots + \beta_{q-1} z^{-(q-1)} + \beta_q z^{-q}, \text{ where } p \geq q \tag{3-5b}$$

Multiplying Eq. (3-4) by $\alpha\left(z^{-1}\right)$ and expanding the DT transfer function yields:

$$\alpha\left(z^{-1}\right) y(t_k) = \beta\left(z^{-1}\right) u(t_k) + \alpha\left(z^{-1}\right) e(t_k) \tag{3-6a}$$

$$\left[1 + \alpha_1 z^{-1} + \ldots + \alpha_p z^{-p}\right] y(t_k) = \left[\beta_0 + \beta_1 z^{-1} + \ldots + \beta_q z^{-q}\right] u(t_k) + \alpha\left(z^{-1}\right) e(t_k) \tag{3-6b}$$

$$y(t_k) + \alpha_1 y(t_{k-1}) + \ldots + \alpha_p y(t_{k-p}) = \beta_0 u(t_k) + \beta_1 u(t_{k-1}) + \ldots + \beta_q u(t_{k-q}) + \alpha\left(z^{-1}\right) e(t_k) \tag{3-6c}$$

$$y(t_k) = -\alpha_1 y(t_{k-1}) - \ldots - \alpha_p y(t_{k-p}) + \beta_0 u(t_k) + \beta_1 u(t_{k-1}) + \ldots + \beta_q u(t_{k-q}) + \alpha\left(z^{-1}\right) e(t_k) \tag{3-6d}$$

The resulting equation in Eq. (3-6d) has transformed the parameter estimation for a DT LTI model into a linear regression problem. Here, the dependent variable is the k^{th} instant of the measured output, i.e. $y(t_k)$, while the independents are the past of the output and the current and past samples of the input, assuming no delay.

Based on the linear regression theory, a vector notation of the one-step-ahead predictor of the DT model in Eq. (3-6d) is (Ljung, 1999; Isermann, 2011):

$$\hat{y}\left(t_k \mid \hat{\boldsymbol{\rho}}_d\right) = \boldsymbol{\psi}(t_k)\hat{\boldsymbol{\rho}}_d \tag{3-7}$$

where $\boldsymbol{\psi}(t_k)$ is the DT regressor $\boldsymbol{\psi}(t_k) = \left[-y(t_{k-1}),\ldots,-y(t_{k-p}),u(t_k),\ldots,u(t_{k-q})\right]$;

$\hat{\boldsymbol{\rho}}_d$ is the DT estimated parameter vector $\hat{\boldsymbol{\rho}}_d = \left[\hat{\alpha}_1,...,\hat{\alpha}_p,\hat{\beta}_0,...,\hat{\beta}_q\right]^T$;

$\hat{y}\left(t_k \mid \hat{\boldsymbol{\rho}}_d\right)$ is the one-step-ahead prediction at the k^{th} instant, according to the estimation results of $\beta\left(z^{-1}\right)$ and $\alpha\left(z^{-1}\right)$, which is essentially $\hat{\boldsymbol{\rho}}_d$.

As every element in the regressor vector can be extracted from the available measurements, the parameter estimation problem of this DT regression model is now ready to be solved by any optimisation strategy. The least squares (LS) algorithm and its variants have been previously developed and applied for this purpose (Young, 1985; Isermann, 2006). The result will be the estimated parameter vector ($\hat{\boldsymbol{\rho}}_d$) that minimises the sum of the squared prediction error, i.e. $\sum [y(t_k) - \hat{y}(t_k|\hat{\boldsymbol{\rho}}_d)]^2$.

Because the LS algorithm is one of convex optimisation, a closed-form solution can be obtained by solving the "normal equation", which can be expressed as follows (Åström and Eykhoff, 1971; Söderström and Stoica, 2002):

$$\hat{\boldsymbol{\rho}}_d = \left[\sum_{k=1}^{N} \boldsymbol{\psi}^T(t_k)\boldsymbol{\psi}(t_k)\right]^{-1}\left[\sum_{k=1}^{N} \boldsymbol{\psi}^T(t_k)\, y(t_k)\right] \tag{3-8}$$

Once the DT model estimation is completed, the result will be converted to its original CT model structure to facilitate the FDD capabilities. In this final process, the system parameters will be computed from the CT model parameters. If the parameter changes are noticeable, then they can be inferred as fault developments in the system.

Therefore, the essential procedure required for indirect parameter estimation falls into *the two-stage domain conversion between CT and DT models* (Unbehauen and Rao, 1990; Garnier, 2015). This requirement will later be evaluated to determine if the indirect parameter estimation is suitable for railway turnout FDD application.

3.2.2 Direct parameter estimation principle

In this case, the algorithms will try to directly apply the same linear regression solution to estimate the CT model parameters (Young, 1981; Garnier and Wang, 2008; Gilson, 2015). Although the measurements are still sampled, recent data acquisition technology is now capable of obtaining data at high sampling rates, enabling sampled data to be close to continuous (Padilla, 2017). Once the CT LTI model in Eq. (3-2) is multiplied by $A(s)$ and the CT transfer function is expanded, it will yield:

$$A(s)y(t) = B(s)u(t) + A(s)n(t) \tag{3-9a}$$

$$\left[s^n + a_1 s^{n-1} + ... + a_n\right]y(t) = \left[b_0 s^m + b_1 s^{m-1} + ... + b_m\right]u(t) + A(s)n(t) \tag{3-9b}$$

$$s^n y(t) + a_1 s^{n-1} y(t) + ... + a_n y(t) = b_0 s^m u(t) + b_1 s^{m-1} u(t) + ... + b_m u(t) + A(s)n(t) \tag{3-9c}$$

$$s^n y(t) = -a_1 s^{n-1} y(t) - ... - a_n y(t) + b_0 s^m u(t) + b_1 s^{m-1} u(t) + ... + b_m u(t) + A(s)n(t) \tag{3-9d}$$

Assuming that all initial conditions are equal to zeros, Eq. (3-9d) can be rewritten as:

$$y^{(n)}(t) = -a_1 y^{(n-1)}(t) - ... - a_n y(t) + b_0 u^{(m)}(t) + b_1 u^{(m-1)}(t) + ... + b_m u(t) + A(s)n(t) \tag{3-9e}$$

This procedure has similarly transformed the parameter estimation of the CT LTI model into a linear regression problem (Maruta and Sugie, 2013). However, in this scenario, the dependent variable is the n^{th} derivative of the measured output, and the independents are the lower derivatives of input and output. Referring to Eq. (3-7), the n^{th} derivative predictor can be similarly calculated as follows:

$$\hat{y}^{(n)}(t \mid \hat{\boldsymbol{\rho}}_c) = \boldsymbol{\mu}(t)\hat{\boldsymbol{\rho}}_c \tag{3-10}$$

where $\boldsymbol{\mu}(t)$ is the CT regressor $\boldsymbol{\mu}(t) = \left[-y^{(n-1)}(t),...,-y(t),u^{(m)}(t),...,u(t)\right]$;

$\hat{\boldsymbol{\rho}}_c$ is the CT estimated parameter vector $\hat{\boldsymbol{\rho}}_c = \left[\hat{a}_1,...,\hat{a}_n,\hat{b}_0,...,\hat{b}_m\right]^T$;

$\hat{y}^{(n)}\left(t \mid \hat{\boldsymbol{\rho}}_c\right)$ is the n^{th} derivative prediction based on $\hat{\boldsymbol{\rho}}_c$.

To enable the optimisation for direct parameter estimation, the sum of squared prediction errors must be computable. It requires all CT variables in the $\boldsymbol{\mu}(t)$ and the n^{th} derivative to be sampled first, as shown in Figure 3-1.

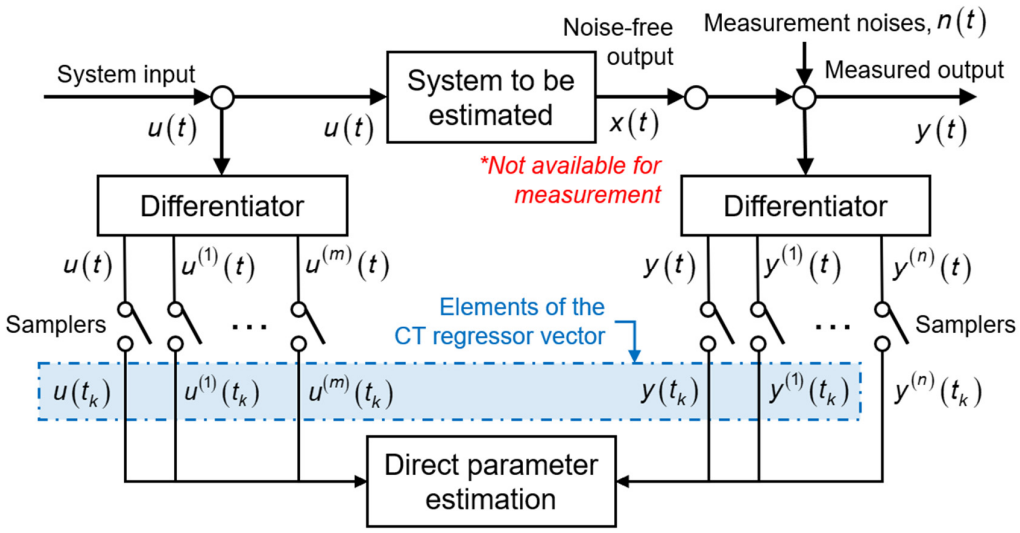


Figure 3-1 Process of obtaining sampled CT variables for direct parameter estimation

Then, a CT normal equation can be formulated as follows:

$$\hat{\boldsymbol{\rho}}_c = \left[\sum_{k=1}^{N} \boldsymbol{\mu}^T(t_k)\boldsymbol{\mu}(t_k)\right]^{-1}\left[\sum_{k=1}^{N} \boldsymbol{\mu}^T(t_k)y^{(n)}(t_k)\right] \tag{3-11}$$

where $y^{(n)}(t_k)$ is the sampled n^{th} derivative of the system output;

$\boldsymbol{\mu}(t_k)$ is the sampled CT regressor = $\left[-y^{(n-1)}(t_k),...,-y(t_k),u^{(m)}(t_k),...,u(t_k)\right]$.

Because domain conversion is no longer required, the system parameters can be directly calculated from the CT model parameter estimates. Then, the evaluation based on the system parameter changes can be performed as in the indirect approach. Since the derivatives are usually not available for measurements, the requirement for direct parameter estimation falls into *the efficient approximation of these derivatives*.

3.2.3 Comparison between indirect and direct parameter estimation

First, indirect parameter estimation for FDD purposes requires two domain conversion stages. Nowadays, these can be conducted by using two function routines, called "`d2c`" and "`c2d`", available in MATLAB® (Mathworks®, 2023). Considering the relationship between the CT *s*-domain and DT *z*-domain transfer functions, the roots of their denominator polynomial are related as follows:

$$p_d = e^{p_c T_s} \tag{3-12}$$

where p_c is one of the poles from the original CT model;

p_d is the associated DT pole resulting from the domain conversion;

T_s is the selected fixed sampling interval.

This equation indicates that *the sampling interval influences the transformation of poles*. When the sampling rate is high, the DT poles will be in proximity to the unit circle boundary, which defines the stability in the *z*-domain, risking the DT equivalent model being unstable. Meanwhile, a sampling rate that is too low will not be able to capture the system's fast dynamics. These demonstrate that appropriately selecting a sampling rate becomes critical for indirect parameter estimation (Jiang et al., 2017).

Additionally, when the algorithm is applied to a system with a degree of freedom higher than one, there are several poles to be estimated. If a second-order system is also "stiff", such that two poles are quite separated, the selected sampling rate will only be suitable to estimate one pole (Garnier, Mensler, and Richard, 2003). Because the mechanical transfer function of railway turnouts to be derived later is a second-order model, combined with the influences of sampling rate in the domain conversion stages, *indirect parameter estimation is deemed not suitable for this particular application*.

In contrast, direct parameter estimation only requires an efficient approximation of the system variable's derivatives. Many techniques have been proposed in the past.

For example, linear filters, such as the generalised Poisson moment (GPM) and state-variable filter (SVF) variants (Rao and Unbehauen, 2006), have been employed to reduce the high-frequency truncation errors caused by the numerical differentiation. Linear integral filter (LIF) techniques have also been applied to avoid differentiation by performing many integrations instead (Garnier, Mensler, and Richard, 2003).

Although solutions for derivative approximation have been proposed, solving the normal equation in Eq. (3-11) will still result in ***statistically inefficient*** parameter estimates. In other words, they are asymptotically biased with a significant estimation error covariance (Ljung, 1999; Young, 2012). These undesirable characteristics exist primarily because of the measurement noises, which are latent in the regressor vector.

Therefore, in general, direct parameter estimation will be an effective solution if it can (1) efficiently approximate the variable's derivatives from sampled data, (2) implement a filtering process to reduce the truncation errors, and (3) suppress the asymptotic bias and covariance of the parameter estimation errors adequately.

The Simplified Refined Instrumental Variable for Continuous Systems (SRIVC) algorithm has been demonstrated to solve all these problems successfully in previous applications (Young, 2015; Janot et al., 2017; Padilla et al., 2019; Dutta et al., 2022). As a result, its adaptation for railway turnout FDD will be the main interest in this thesis. In the following section, the SRIVC algorithm is described in greater detail.

3.3 The SRIVC parameter estimation algorithm

The SRIVC algorithm has been proven to be capable of producing statistically efficient results regardless of the stochastic characteristics of the measurement noises (Pan et al., 2020a; 2020b). This is achievable as the algorithm implements iteratively updated instrumental variables (IVs) and pre-filters, which are elaborated as follows.

3.3.1 The solution to asymptotic bias: instrumental variables (IVs)

The bias is caused by the multiplication product between the regressor vector and its transpose, both of which contain measurement noises. The idea to reduce this asymptotic bias is, therefore, to remove these self-correlations of noises. First, the CT normal equation derived from the LS algorithm is repeated here for convenience:

$$\hat{\boldsymbol{\rho}}_c = \left[\sum_{k=1}^{N} \boldsymbol{\mu}^T(t_k)\boldsymbol{\mu}(t_k)\right]^{-1}\left[\sum_{k=1}^{N} \boldsymbol{\mu}^T(t_k)y^{(n)}(t_k)\right] \tag{3-11}$$

Then, the transposed regressor vector is replaced with one that does not correlate with latent measurement noises as follows (Söderström, 2011):

$$\hat{\boldsymbol{\rho}}_{IV} = \left[\sum_{k=1}^{N} \boldsymbol{\eta}^T(t_k)\boldsymbol{\mu}(t_k)\right]^{-1}\left[\sum_{k=1}^{N} \boldsymbol{\eta}^T(t_k)y^{(n)}(t_k)\right] \tag{3-13}$$

where $\boldsymbol{\eta}(t_k)$ is the sampled CT instrument vector that replaces all measured output elements with instrumental variables, i.e. $z(t_k)$, which are uncorrelated with noises, resulting in $\left[-z^{(n-1)}(t_k),...,-z(t_k),u^{(m)}(t_k),...,u(t_k)\right]$.

For Eq. (3-13) to be solvable, the matrix in the left bracket must still be full-rank. In other words, $\sum \boldsymbol{\eta}^T(t_k)\boldsymbol{\mu}(t_k)$ is still invertible. This results in the first requirement that IVs should <u>*not be correlated with the existing measurement noises but correlated with the inputs and the theoretical noise-free outputs*</u>, i.e. $u(t_k)$ and $x(t_k)$.

As a consequence, the SRIVC algorithm proposes an automated IV generation by adopting the approximation of "true" theoretical noise-free output $x(t_k)$, i.e. $\hat{x}(t_k)$, as its IVs, such that the instrument vector used in the SRIVC algorithm is:

$$\boldsymbol{\eta}(t_k) = \left[-\hat{x}^{(n-1)}(t_k),...,-\hat{x}(t_k),u^{(m)}(t_k),...,u(t_k)\right] \tag{3-14}$$

where $\hat{x}(t_k)$ is the sampled approximation of the "true" theoretical noise-free output, which can be calculated from $\hat{x}(t,\hat{\boldsymbol{\rho}}_{IV}) = \left[\hat{B}(s)/\hat{A}(s)\right]u(t)$.

This selection ensures that the IVs will also be correlated with the deterministic system input since the estimated CT model's lag is accounted for. The degree of its correlation to the "true" theoretical noise-free output will depend on the quality of the CT parameter estimates. Nevertheless, **they are paradoxically the objective of the parameter estimation**. As a result, the SRIVC algorithm employs an iteration process where the estimate $\hat{x}\left(t,\hat{\boldsymbol{\rho}}_{IV}\right)$ is iteratively updated as follows (Young, 2015):

$$\hat{x}^{\langle j\rangle}\left(t,\hat{\boldsymbol{\rho}}_{IV}^{\langle j-1\rangle}\right)=\frac{\hat{B}^{\langle j-1\rangle}(s)}{\hat{A}^{\langle j-1\rangle}(s)}u(t) \tag{3-15}$$

where $\hat{x}^{\langle j\rangle}\left(t,\hat{\boldsymbol{\rho}}_{IV}^{\langle j-1\rangle}\right)$ is the estimated theoretical noise-free output at the j^{th} iteration;

$\hat{B}^{\langle j-1\rangle}(s)$ and $\hat{A}^{\langle j-1\rangle}(s)$ are the estimates of $B(s)$ and $A(s)$ polynomials from the previous round of iteration (i.e. from the $(j-1)^{\text{th}}$ iteration).

Since the normal equation in Eq. (3-13) is still a closed-form optimisation, it is still guaranteed to find the global minimum. As a result, the iteration process will converge the parameter estimates to their "true" values. If the number of samples approaches infinity, it has been proven that the estimated noise-free output will also be converged to its "true" value, i.e. $\hat{x}(t)\to x(t)$ (Åström and Eykhoff, 1971).

Although infinite samples are not practical, this statistical property suggests that the more samples being used in Eq. (3-13), the more accurate the estimation results will be. Suppose the total duration of the measurement is fixed, like the turnout throw time. In that case, a higher sampling rate can be implemented to obtain more samples.

3.3.2 Reduction of numerical differentiation errors by using pre-filters

A numerical differentiation process is mandatory for CT model estimation. One of its kind is the backward Euler method, which can be expressed as follows:

$$\frac{df}{dt} \approx \frac{\Delta f}{\Delta t} = \frac{f(t_k) - f(t_{k-1})}{\Delta t} \tag{3-16}$$

where Δt is the sampling interval (i.e. a reciprocal of the sampling rate).

In reality, there are always measurement noises that cause the true values and their measurements to be different. Approximation of their derivative by this numerical process will amplify these differences by a factor of $1/\Delta t$ (which equals the sampling frequency). Since the sampling interval should be as close to zero as possible to obtain statistically optimal estimates, *low-pass filters have become necessary*.

Several algorithms have proposed different pre-filters, such as SVF and GPM. However, these require the users to define their hyperparameters, which usually come in the form of the cut-off frequency. Since it is one of the system characteristics that are initially unknown, the users must have sufficient domain knowledge to be able to select this value appropriately (Garnier, Mensler, and Richard, 2003).

The SRIVC algorithm takes another approach by attempting to automate the pre-filter selection by an iteration process. First, Eq. (3-9a) is repeated for convenience:

$$A(s)y(t) = B(s)u(t) + A(s)n(t) \tag{3-9a}$$

Then, the entire equation is divided by the denominator of the CT transfer function:

$$\left(\frac{1}{A(s)}\right)A(s)y(t) = \left(\frac{1}{A(s)}\right)B(s)u(t) + \left(\frac{1}{A(s)}\right)A(s)n(t) \tag{3-17a}$$

Instead of cancelling the $A(s)$, the filtered variables are defined as follows:

$$y_f(t) = \left(\frac{1}{A(s)}\right)y(t) \text{ and } u_f(t) = \left(\frac{1}{A(s)}\right)u(t) \tag{3-17b}$$

Substituting the expressions in Eq. (3-17b) into Eq. (3-17a) obtains:

$$A(s)y_f(t) = B(s)u_f(t) + n(t) \tag{3-17c}$$

Assuming that all initial conditions are zeros, this filtered CT model can be expanded to form the following linear regression equation:

$$\left[s^n + a_1 s^{n-1} + \ldots + a_n\right] y_f(t) = \left[b_0 s^m + b_1 s^{m-1} + \ldots + b_m\right] u_f(t) + n(t) \tag{3-17d}$$

$$s^n y_f(t) = -a_1 s^{n-1} y_f(t) - \ldots - a_n y_f(t) + b_0 s^m u_f(t) + b_1 s^{m-1} u_f(t) + \ldots + b_m u_f(t) + n(t) \tag{3-17e}$$

$$y_f^{(n)}(t) = -a_1 y_f^{(n-1)}(t) - \ldots - a_n y_f(t) + b_0 u_f^{(m)}(t) + b_1 u_f^{(m-1)}(t) + \ldots + b_m u_f(t) + n(t) \tag{3-17f}$$

Lastly, the normal equation is derived from Eq. (3-17f), and IVs are applied by replacing the transposed regressor vector with the transposed instrument vector, resulting in:

$$\hat{\boldsymbol{\rho}}_{SRIVC} = \left[\sum_{k=1}^{N} \boldsymbol{\eta}_f^T(t_k)\boldsymbol{\mu}_f(t_k)\right]^{-1}\left[\sum_{k=1}^{N} \boldsymbol{\eta}_f^T(t_k) y_f^{(n)}(t_k)\right] \tag{3-18}$$

where $\boldsymbol{\eta}_f(t_k)$ is the sampled filtered CT instrument vector where all elements in $\boldsymbol{\eta}(t)$ are filtered by the CT pre-filter $\left[1/A(s)\right]$;

$\boldsymbol{\mu}_f(t_k)$ is the sampled filtered CT regressor vector where all elements in $\boldsymbol{\mu}(t)$ are filtered by the CT pre-filter $\left[1/A(s)\right]$;

$y_f^{(n)}(t_k)$ is the sampled of the n^{th} derivative approximation of the measured output, i.e. $y^{(n)}(t)$, that is filtered by the CT pre-filter $\left[1/A(s)\right]$.

This approach eliminates the manual process of pre-filter selection by iteratively updating it based on the previous estimation results. To conclude this methodology, the following section summarises the complete procedure of the SRIVC algorithm that is proposed to be an effective CT model estimation solution.

3.3.3 Procedural steps of the SRIVC algorithm

All of the procedures of the SRIVC algorithm can be described as follows:

Step 1: An initialisation process is required to calculate the IVs and determine the pre-filter for the first round of iterations. In this thesis, the SRIV algorithm is used to provide

the initial results (Young, 1985). In cases where the resulting DT model is unstable, which may happen especially in high sampling rate scenarios, it will be made stable first by using the "`polystab`" function routine in MATLAB®. Then, the stable DT model will be converted into an initial CT transfer function via the "`d2c`" routine.

Step 2: After the initialisation, the iteration process will be commenced. Until it reaches the maximum number of iterations or the parameter estimates are all converged, i.e. when $\left\| \hat{\boldsymbol{\rho}}^{\langle j-1 \rangle} - \hat{\boldsymbol{\rho}}^{\langle j \rangle} \right\| < \delta$ where δ is a small arbitrary number, each round will use the result from the previous iteration to estimate the theoretical noise-free outputs and use them as the IVs for the current iteration, as follows:

$$\hat{x}^{\langle j \rangle}(t) = \frac{\hat{B}^{\langle j-1 \rangle}(s)}{\hat{A}^{\langle j-1 \rangle}(s)} u(t) \tag{3-19}$$

where $\hat{x}^{\langle j \rangle}(t)$ is the estimated theoretical noise-free output (the IV) of the j^{th} iteration;

$\hat{B}^{\langle j-1 \rangle}(s)$ and $\hat{A}^{\langle j-1 \rangle}(s)$ are the estimated numerator and denominator terms from the previous iteration round, indicated by the superscript '$\langle j-1 \rangle$'.

NOTE: If the estimated CT transfer function from the previous round is unstable (i.e. the real part of any pole is positive), the unstable poles must be reflected to the left-hand side of the *s*-plane first to ensure their stability before use in Eq. (3-19).

Then, the measured inputs, outputs, and IVs are all filtered by $\left[1 / \hat{A}^{\langle j-1 \rangle}(s) \right]$. At the end of each iteration, the parameter estimates can be computed as follows:

$$\hat{\boldsymbol{\rho}}_{\text{SRIVC}}^{\langle j \rangle} = \left[\sum_{k=1}^{N} \boldsymbol{\eta}_f^T \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle j-1 \rangle} \right) \boldsymbol{\mu}_f^T \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle j-1 \rangle} \right) \right]^{-1} \left[\sum_{k=1}^{N} \boldsymbol{\eta}_f^T \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle j-1 \rangle} \right) y_f^{(n)} \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle j-1 \rangle} \right) \right] \tag{3-20}$$

where $\hat{\boldsymbol{\rho}}_{\text{SRIVC}}^{\langle j \rangle}$ is the SRIVC estimated parameter vector of the j^{th} iteration.

Every component in this normal equation is dependent on the estimates from the previous iteration, which is indicated by $\hat{\boldsymbol{\rho}}^{\langle j-1 \rangle}$. Therefore, for the first iteration, both IVs and the pre-filter will be formulated from the initialisation result, i.e. $\hat{\boldsymbol{\rho}}^{\langle 0 \rangle}$ (Step 1).

Step 3: After the iteration is completed, the final estimation results will be evaluated based on their statistical properties. There are two criteria commonly used for SRIVC parameter estimation (Young, 2012; Garnier, 2015; Pascu et al., 2019).

The first is the coefficient of determination (R_T^2) which quantifies the data fitting performance. This criterion is defined as the portion of the measured output's variance that the model output can represent, which can be calculated as follows:

$$R_T^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2} \tag{3-21}$$

where R_T^2 is the coefficient of determination;

σ_ε^2 is the variance of the model prediction errors, i.e. $\varepsilon(t_k) = y(t_k) - \hat{x}^{\langle f \rangle}(t_k)$, where $\langle f \rangle$ indicates the final round of iteration;

σ_y^2 is the variance of the measured system output.

The ideal scenario is when there is no prediction error, resulting in a unity R_T^2.

The second criterion is derived from an assumption that the model prediction error has "white noise" properties (i.e. zero mean, constant variance, and negligible covariances with the measured inputs). If this assumption holds, the covariance matrix of the parameter estimation error can be approximated as follows (Young, 2012):

$$COV_{\hat{\boldsymbol{\rho}}} = \sigma_\varepsilon^2 \left[\sum_{k=1}^{N} \boldsymbol{\mu}_f^T \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle f \rangle} \right) \boldsymbol{\mu}_f^T \left(t_k \mid \hat{\boldsymbol{\rho}}^{\langle f \rangle} \right) \right]^{-1} \tag{3-22}$$

A high percentage of the coefficient of determination, small diagonal entries in the covariance error matrix related to their corresponding parameter estimates, and a

negligible amount of all off-diagonal entries are the desired properties of the estimation results. Collectively, they indicate the statistical efficiency of the parameter estimates.

Since the measurement noises are stochastic, all parameter estimates will be as well. Suppose the noises are assumed to be normally distributed. In that case, the mean value of the estimate will equal the element in the vector $\hat{\boldsymbol{\rho}}^{\langle f \rangle}$ with a standard deviation equal to the square root of the corresponding diagonal entries in $COV_{\hat{\boldsymbol{\rho}}}$.

All these procedural steps of the SRIVC algorithm can be graphically illustrated in Figure 3-2. In the next section, implementation of the SRIVC algorithm to estimate railway electro-mechanical turnout parameters will commence.

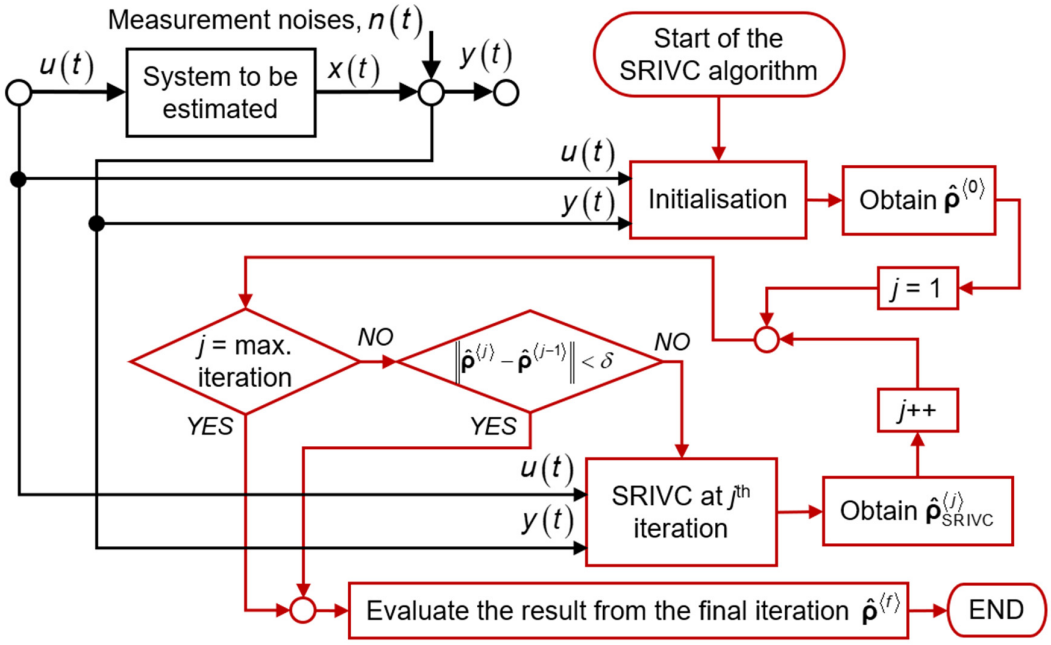


Figure 3-2 High-level flowchart of SRIVC algorithm procedures

3.4 Simulation model of DC electro-mechanical turnouts

To understand how the SRIVC algorithm can be applied and to create data for assessing its performance, in this section, a DC electro-mechanical turnout simulation model is developed. There are two reasons for choosing this specific type: (1) it has similar characteristics compared to an HW55S turnout available as a test system in the

laboratory (which will be described in detail in Chapter 5), and (2) it is commonly used in UK railways (Oyebande and Renfrew, 2002; Matei et al., 2015; Jackson, 2023).

To conduct a preliminary verification of the SRIVC algorithm, the switch rail's bending dynamics are initially modelled as a mass-spring-damper load for simplicity, which has been employed in many previous research studies (Roberts et al., 2002; Zattoni, 2006; Bai, 2010; Hamadache et al., 2019b; Dutta et al., 2022). A more realistic representation is then presented and discussed in Chapter 4.

The actuation element of this model consists of a permanent magnet DC motor, a gearbox, a ball screw to convert a rotational motion into a translational one, and a mass-spring-damper load to represent the switch rail's dynamics. Furthermore, every coupling in the model is assumed to be rigidly connected and stiff (Matei et al., 2015; Li et al., 2022). A schematic of this simulation is illustrated in Figure 3-3.

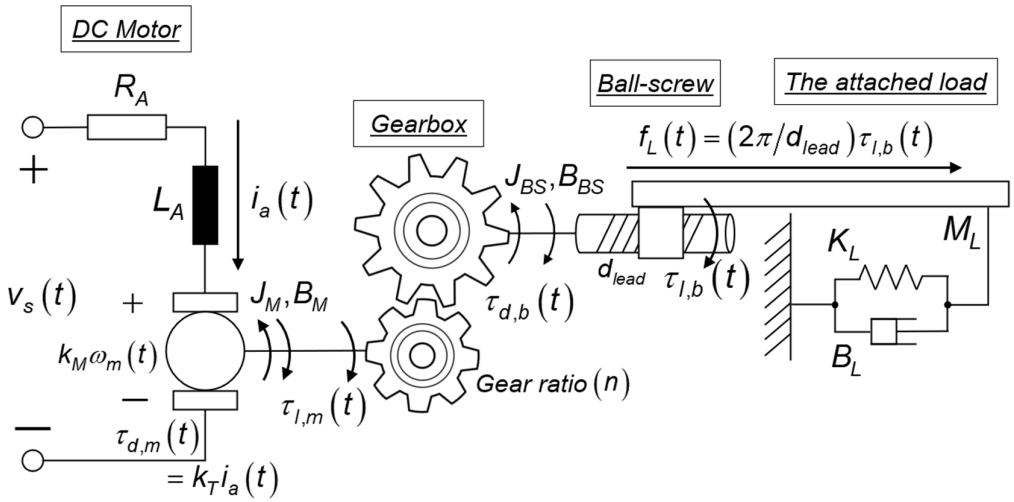


Figure 3-3 Schematic diagram of a permanent magnet DC simulation model that represents the linear actuation mechanism of a DC railway turnout

Applying Kirchhoff's Voltage Law (KVL) to the DC motor, its commonly known electrical differential equation is formulated. Meanwhile, using Newton's second law of rotation shall appropriately explain the DC motor's mechanical dynamics. Additionally, since a permanent magnet DC motor is considered in this simulation model, there are

three more relationships that link these two ordinary differential equations (ODEs). As a result, all of the related equations and relationships are:

$$v_s(t) = L_A \frac{di_a(t)}{dt} + R_A i_a(t) + e_a(t) \tag{3-23a}$$

$$\tau_{d,m}(t) = J_M \frac{d\omega_m(t)}{dt} + B_M \omega_m(t) + \tau_{l,m}(t) \tag{3-23b}$$

$$e_a(t) = k_M \omega_m(t),\ \tau_{d,m}(t) = k_T i_a(t), \text{ and } k_M = k_T \tag{3-23c}$$

where $v_s(t)$ is the input voltage supplied across the DC motor's terminals;

$i_a(t)$ is the current passing through the DC motor's armature coil;

$e_a(t)$ is the back electromotive force (EMF) generated from the rotation of the armature coil within the static magnetic field;

$\omega_m(t)$ is the angular speed of the DC motor shaft;

$\tau_{d,m}(t)$ is the driving torque produced by the DC motor;

$\tau_{l,m}(t)$ is the required load torque as seen by the DC motor;

R_A is the armature resistance, the unit is Ω;

L_A is the armature inductance, the unit is H;

k_M is the voltage constant of the DC motor, the unit is V/(rad/s);

k_T is the torque constant of the DC motor, the unit is Nm/A;

J_M is the moment of inertia of the motor shaft, the unit is kg·m^2;

B_M is the rotational viscous damping of the motor shaft, the unit is Nm·s/rad.

The motor shaft is then coupled with the input side of the reduction gearbox. Assuming that this gearbox is ideal (i.e. no inertia, elasticity, damping, and backlash), the relationships between its input and output side can be formulated as follows:

$$\tau_{d,b}(t) = n\tau_{l,m}(t) \tag{3-24a}$$

$$\omega_b(t) = (1/n)\omega_m(t) \tag{3-24b}$$

where n is the gear reduction ratio of this gearbox (n:1 = output side : input side);

$\tau_{d,b}(t)$ is the torque at the output side of the gearbox that will be coupled to drive the ball screw;

$\omega_b(t)$ is the angular speed of the ball screw shaft, which is connected to the output side of the gearbox.

The following transmission stage is the ball screw, which converts the driving torque to an output actuating force. Similar to the mechanical differential equation of the DC motor, Newton's second law of rotation is applicable here as well. Therefore, the related equations that describe the ball screw's dynamics are as follows:

$$\tau_{d,b}(t) = J_{BS}\frac{d\omega_b(t)}{dt} + B_{BS}\omega_b(t) + \tau_{l,b}(t) \tag{3-25a}$$

$$\tau_{l,b}(t) = \frac{d_{lead}}{2\pi} f_L(t) \tag{3-25b}$$

where $\tau_{l,b}(t)$ is the required load torque as seen by the ball screw;

$f_L(t)$ is the converted actuating force from the ball screw to move the load;

J_{BS} is the moment of inertia of the ball screw, the unit is kg·m^2;

B_{BS} is the rotational viscous damping of the ball screw, the unit is Nm·s/rad;

d_{lead} is the moving distance of the ball screw's table attached to the load when the ball screw's shaft rotates one revolution, the unit is m.

Finally, since a load of this simulation is modelled as a mass-spring-damper assembly, the second law of Newton's linear motion can be implemented to describe its mechanical response. As a result, the load's ODE can be expressed as follows:

$$f_L(t) = M_L\frac{d^2x_{load}(t)}{dt} + B_L\frac{dx_{load}(t)}{dt} + K_L x_{load}(t) \tag{3-26}$$

where $x_{load}(t)$ is the displacement of the load that is being moved by the DC actuator;

M_L is the mass of the attached load, the unit is kg;

B_L is the translational viscous damping of the attached load, the unit is N·s/m;

K_L is the stiffness of the attached load, the unit is N/m.

Then, all ODEs and relationships from Eqs. (3-23) to (3-26) are implemented in Simulink® to create the DC turnout simulation model, as shown in Figure 3-4.

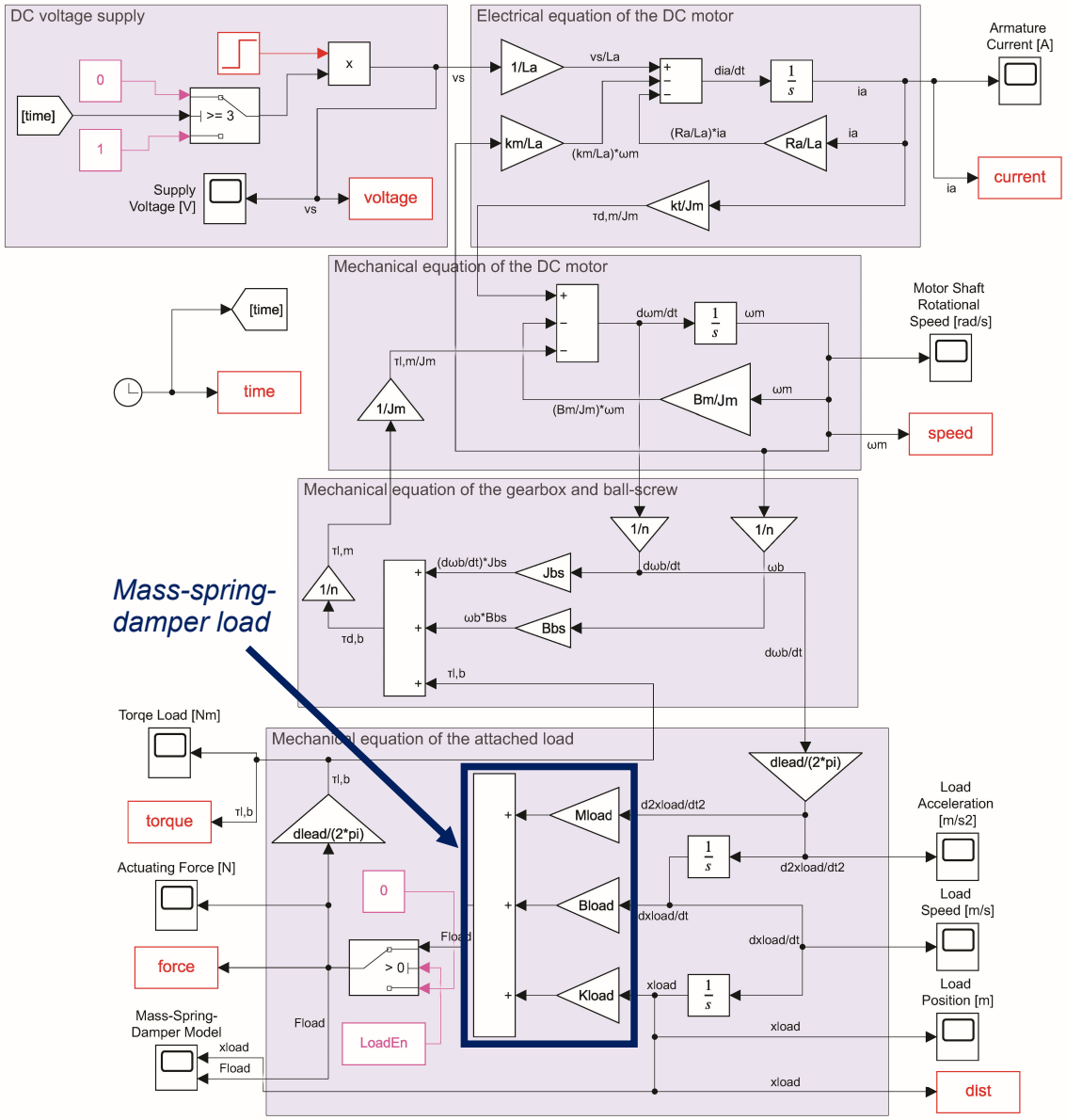


Figure 3-4 Permanent magnet DC turnout simulation model implemented in Simulink®

This turnout simulation model is used to generate the measurements required for parameter estimation, where the statistical performance of several algorithms is assessed and compared. The following section demonstrates the application of the SRIVC algorithm to the equations and relationships that govern the turnout dynamics.

3.5 Application of the SRIVC algorithm to DC railway turnouts

To begin with, the first model considered is the electrical transfer function of DC turnout systems. First, Eq. (3-23) is repeated here for convenience:

$$v_s(t) = L_A \frac{di_a(t)}{dt} + R_A i_a(t) + e_a(t) \tag{3-23a}$$

$$\tau_{d,m}(t) = J_M \frac{d\omega_m(t)}{dt} + B_M \omega_m(t) + \tau_{l,m}(t) \tag{3-23b}$$

$$e_a(t) = k_M \omega_m(t),\ \tau_{d,m}(t) = k_T i_a(t), \text{ and } k_M = k_T \tag{3-23c}$$

Due to the physical structure of DC motors, their back EMF is not measurable. However, Eq. (3-23c) indicates that there is a relationship between the back EMF and the rotational motor shaft speed that is measurable. This will facilitate the formation of the DC turnout's electrical transfer function, which starts from the following ODE:

$$v_s(t) = L_A \frac{di_a(t)}{dt} + R_A i_a(t) + k_M \omega_m(t) \tag{3-27}$$

The dependence of "time" in each variable is included to differentiate them from their Laplace transforms. Assuming that the initial condition of motor current is zero, the Laplace transformation of this electrical ODE yields:

$$V_s(s) = L_A s I_a(s) + R_A I_a(s) + k_M \Omega_m(s) \tag{3-28a}$$

$$(sL_A + R_A) I_a(s) = V_s(s) - k_M \Omega_m(s) \tag{3-28b}$$

$$I_a(s) = \left(\frac{1}{sL_A + R_A}\right) V_s(s) + \left(\frac{-k_M}{sL_A + R_A}\right) \Omega_m(s) \tag{3-29}$$

This transfer function model is not of a SISO type since it requires two inputs, resulting in a multiple-input single-output (MISO) model instead. Here, the variables needed for parameter estimation include the input voltage $v_s(t)$, the armature current $i_a(t)$, and the rotational speed of the motor shaft $\omega_m(t)$.

Considering the measurement process of this electrical transfer function by adding a Laplace transform of measurement "white noise" term, $N(s)$, as follows:

$$I_a(s) = \left(\frac{1}{sL_A + R_A}\right)V_s(s) + \left(\frac{-k_M}{sL_A + R_A}\right)\Omega_m(s) + N(s) \tag{3-30}$$

Comparing this equation to Eq. (3-17a), the pre-filter to be employed in this electrical transfer function is, therefore, $[1/(sL_A + R_A)]$. Thus, the entire Eq. (3-30) is multiplied by such a pre-filter to yield the following expression:

$$\begin{aligned}\left[\frac{1}{sL_A + R_A}\right]I_a(s) = &\left[\frac{1}{sL_A + R_A}\right]\left(\frac{1}{sL_A + R_A}\right)V_s(s) \\ &+ \left[\frac{1}{sL_A + R_A}\right]\left(\frac{-k_M}{sL_A + R_A}\right)\Omega_m(s) + \left[\frac{1}{sL_A + R_A}\right]N(s)\end{aligned} \tag{3-31}$$

Then, defining $I_{a,f}(s)$ as $[1/(sL_A + R_A)]I_a(s)$ and all other variables in the same manner. This substitution results in the following equation:

$$I_{a,f}(s) = \left[\frac{1}{sL_A + R_A}\right]V_{s,f}(s) + \left[\frac{-k_M}{sL_A + R_A}\right]\Omega_{m,f}(s) + \left[\frac{1}{sL_A + R_A}\right]N(s) \tag{3-32}$$

This equation is rearranged into the linear regression format as follows:

$$(sL_A + R_A)I_{a,f}(s) = [1]V_{s,f}(s) + [-k_M]\Omega_{m,f}(s) + N(s) \tag{3-33a}$$

$$sI_{a,f}(s) = -\left(\frac{R_A}{L_A}\right)I_{a,f}(s) + \left(\frac{1}{L_A}\right)V_{s,f}(s) + \left(\frac{-k_M}{L_A}\right)\Omega_{m,f}(s) + N(s) \tag{3-33b}$$

Eq. (3-33b) is transformed to yield the following time-domain function:

$$\frac{di_{a,f}(t)}{dt} = -\left(\frac{R_A}{L_A}\right)i_{a,f}(t) + \left(\frac{1}{L_A}\right)v_{s,f}(t) + \left(\frac{-k_M}{L_A}\right)\omega_{m,f}(t) + n(t) \tag{3-33c}$$

Eq. (3-33c) can be expressed in the sampled vector notation as follows:

$$i_{a,f}^{(1)}(t_k) = \boldsymbol{\mu}_{\text{elec,f}}(t_k)\boldsymbol{\rho}_{\text{elec}} + n(t_k) \tag{3-34}$$

where $\boldsymbol{\mu}_{\text{elec,f}}(t_k)$ is a sampled vector containing filtered electrical regressor elements

$$\boldsymbol{\mu}_{\text{elec,f}}(t_k) = \left[-i_{a,f}(t_k), v_{s,f}(t_k), \omega_{m,f}(t_k)\right];$$

$\boldsymbol{\rho}_{\text{elec}}$ is a vector containing the electrical transfer function parameters:

$$\boldsymbol{\rho}_{\text{elec}} = \left[\left(\frac{R_A}{L_A}\right),\left(\frac{1}{L_A}\right),\left(\frac{-k_M}{L_A}\right)\right]^T = \left[a_{e,1}, b_{e,10}, b_{e,20}\right]^T.$$

Therefore, when the SRIVC algorithm is performed on Eq. (3-34), the elements in the estimated electrical parameter vector, i.e. $\hat{\boldsymbol{\rho}}_{\text{elec}}$, will be related to the electrical parameters of the DC electro-mechanical turnout systems as follows:

$$\hat{L}_A = \left(1/\hat{b}_{e,10}\right),\ \hat{R}_A = \left(\hat{a}_{e,1}/\hat{b}_{e,10}\right), \text{ and } \hat{k}_M = \left(-\hat{b}_{e,20}/\hat{b}_{e,10}\right). \tag{3-35}$$

On the other hand, the formation of the mechanical transfer function requires a consideration of multiple transmission stages. Starting from the final stage, the related equations describing the dynamics of the mass-spring-damper load and ball screw are:

$$\tau_{d,b}(t) = J_{BS}\frac{d\omega_b(t)}{dt} + B_{BS}\omega_b(t) + \tau_{l,b}(t) \text{ and } \tau_{l,b}(t) = \frac{d_{lead}}{2\pi}f_L(t) \tag{3-25}$$

$$f_L(t) = M_L\frac{d^2x_{load}(t)}{dt} + B_L\frac{dx_{load}(t)}{dt} + K_L x_{load}(t) \tag{3-26}$$

These equations are combined to formulate the following ODE:

$$\tau_{d,b}(t) = J_{BS}\frac{d\omega_b(t)}{dt} + B_{BS}\omega_b(t) + \left(\frac{d_{lead}}{2\pi}\right)\left[M_L\frac{d^2x_{load}(t)}{dt} + B_L\frac{dx_{load}(t)}{dt} + K_L x_{load}(t)\right] \tag{3-36a}$$

$$\tau_{d,b}(t) = J_{BS}\frac{d\omega_b(t)}{dt} + B_{BS}\omega_b(t) + \left(\frac{d_{lead}}{2\pi}\right)^2\left[M_L\frac{d^2\theta_b(t)}{dt} + B_L\frac{d\theta_b(t)}{dt} + K_L\theta_b(t)\right] \tag{3-36b}$$

where $\theta_b(t)$ is the angular displacement of the ball screw shaft, which is related to the displacement of load $x_{load}(t)$, as follows: $x_{load}(t) = [d_{lead}/2\pi]\theta_b(t)$.

Since $\theta_b(t) = \int_0^t \omega_b(\tau)d\tau$, assuming that the initial angular displacement equals zero, Eq. (3-36b) can be rewritten as follows:

$$\tau_{d,b}(t)=\left[J_{BS}+\left(\frac{d_{lead}}{2\pi}\right)^2 M_L\right]\frac{d\omega_b(t)}{dt}$$
$$+\left[B_{BS}+\left(\frac{d_{lead}}{2\pi}\right)^2 B_L\right]\omega_b(t)+\left[\left(\frac{d_{lead}}{2\pi}\right)^2 K_L\right]\int_0^t \omega_b(\tau)d\tau \qquad (3\text{-}37)$$

The following stage focuses on the gearbox, which is equivalent to the driving side of the ball screw and the loading side of the DC motor shaft (due to the assumption of an ideal gearbox). First, the gearbox equations are repeated here for convenience:

$$\tau_{d,b}(t)=n\tau_{l,m}(t) \text{ and } \omega_b(t)=(1/n)\omega_m(t) \qquad (3\text{-}24)$$

Transferring all the variables from the output side of the gearbox to the input side by using these equations, the mechanical ODE in Eq. (3-37) can be rewritten as:

$$n\tau_{l,m}(t)=\left(\frac{1}{n}\right)\left[J_{BS}+\left(\frac{d_{lead}}{2\pi}\right)^2 M_L\right]\frac{d\omega_m(t)}{dt}$$
$$+\left(\frac{1}{n}\right)\left[B_{BS}+\left(\frac{d_{lead}}{2\pi}\right)^2 B_L\right]\omega_m(t)+\left(\frac{1}{n}\right)\left[\left(\frac{d_{lead}}{2\pi}\right)^2 K_L\right]\int_0^t \omega_m(\tau)d\tau \qquad (3\text{-}38a)$$

$$\tau_{l,m}(t)=\left[\frac{J_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 M_L\right]\frac{d\omega_m(t)}{dt}$$
$$+\left[\frac{B_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 B_L\right]\omega_m(t)+\left[\left(\frac{d_{lead}}{2\pi n}\right)^2 K_L\right]\int_0^t \omega_m(\tau)d\tau \qquad (3\text{-}38b)$$

Finally, Eq. (3-38b) is combined with the mechanical ODE of the DC motor as follows:

$$k_T i_a(t)=J_M\frac{d\omega_m(t)}{dt}+B_M\omega_m(t)+\left[\frac{J_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 M_L\right]\frac{d\omega_m(t)}{dt}$$
$$+\left[\frac{B_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 B_L\right]\omega_m(t)+\left[\left(\frac{d_{lead}}{2\pi n}\right)^2 K_L\right]\int_0^t \omega_m(\tau)d\tau \qquad (3\text{-}39a)$$

$$k_T i_a(t)=\left[J_M+\frac{J_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 M_L\right]\frac{d\omega_m(t)}{dt}$$
$$+\left[B_M+\frac{B_{BS}}{n^2}+\left(\frac{d_{lead}}{2\pi n}\right)^2 B_L\right]\omega_m(t)+\left[\left(\frac{d_{lead}}{2\pi n}\right)^2 K_L\right]\int_0^t \omega_m(\tau)d\tau \qquad (3\text{-}39b)$$

$$k_T i_a(t) = J_{TO}\frac{d\omega_m(t)}{dt} + B_{TO}\omega_m(t) + K_{TO}\int_0^t \omega_m(\tau)\,d\tau \tag{3-40}$$

where J_{TO} is the turnout's moment of inertia; $J_{TO} = J_M + \frac{J_{BS}}{n^2} + \left(\frac{d_{lead}}{2\pi n}\right)^2 M_L$;

B_{TO} is the turnout's damping effects; $B_{TO} = B_M + \frac{B_{BS}}{n^2} + \left(\frac{d_{lead}}{2\pi n}\right)^2 B_L$;

K_{TO} is the turnout's stiffness, which is assumed only to be contributed by the mass-spring-damper load, not from the actuator; $K_{TO} = \left(\frac{d_{lead}}{2\pi n}\right)^2 K_L$.

The Laplace transformation is conducted on Eq. (3-40) to yield the following transfer function by assuming that the initial motor shaft speed is zero, therefore:

$$k_T I_a(s) = J_{TO}s\Omega_m(s) + B_{TO}\Omega_m(s) + K_{TO}\frac{\Omega_m(s)}{s} \tag{3-41a}$$

$$sk_T I_a(s) = s^2 J_{TO}\Omega_m(s) + sB_{TO}\Omega_m(s) + K_{TO}\Omega_m(s) \tag{3-41b}$$

$$\Omega_m(s) = \left[\frac{k_T}{s^2 J_{TO} + sB_{TO} + K_{TO}}\right] sI_a(s) \tag{3-42}$$

Through these derivations, the resulting Eq. (3-42) is the mechanical transfer function model of the entire turnout system. This model is of a simple SISO type, with the first derivative of the motor current as the input and the rotational speed of the motor shaft as the output. Similarly, considering the measurement process by adding a Laplace transform of the "white noise" to Eq. (3-42) will yield the following equation:

$$\Omega_m(s) = \left[\frac{k_T}{s^2 J_{TO} + sB_{TO} + K_{TO}}\right] sI_a(s) + N(s) \tag{3-43}$$

Comparing this equation to Eq. (3-17a), the pre-filter to be implemented in this mechanical transfer function is, therefore, $\left[1/\left(s^2 J_{TO} + sB_{TO} + K_{TO}\right)\right]$. Then, multiplying the entire Eq. (3-43) with this pre-filter will yield:

$$\left[\frac{1}{s^2J_{TO}+sB_{TO}+K_{TO}}\right]\Omega_m(s)=\left[\frac{1}{s^2J_{TO}+sB_{TO}+K_{TO}}\right]\times$$
$$\left(\frac{k_T}{s^2J_{TO}+sB_{TO}+K_{TO}}\right)sI_a(s)+\left[\frac{1}{s^2J_{TO}+sB_{TO}+K_{TO}}\right]N(s) \qquad (3\text{-}44)$$

Then, defining $\Omega_{m,f}(s)$ as $\left[1/\left(s^2J_{TO}+sB_{TO}+K_{TO}\right)\right]\Omega_m(s)$ and other variables in the same manner. This substitution results in the following equation:

$$\Omega_{m,f}(s)=\left[\frac{k_T}{s^2J_{TO}+sB_{TO}+K_{TO}}\right]sI_{a,f}(s)+\left[\frac{1}{s^2J_{TO}+sB_{TO}+K_{TO}}\right]N(s) \quad (3\text{-}45)$$

This equation is rearranged into the linear regression format as follows:

$$\left(s^2J_{TO}+sB_{TO}+K_{TO}\right)\Omega_{m,f}(s)=\left[k_T\right]sI_{a,f}(s)+N(s) \qquad (3\text{-}46a)$$

$$s^2\Omega_{m,f}(s)=-\left(\frac{B_{TO}}{J_{TO}}\right)s\Omega_{m,f}(s)$$
$$-\left(\frac{K_{TO}}{J_{TO}}\right)\Omega_{m,f}(s)+\left(\frac{k_T}{J_{TO}}\right)sI_{a,f}(s)+N(s) \qquad (3\text{-}46b)$$

Eq. (3-46b) is transformed to yield the following time-domain equation:

$$\frac{d^2\omega_{m,f}(t)}{dt^2}=-\left(\frac{B_{TO}}{J_{TO}}\right)\frac{d\omega_{m,f}(t)}{dt}-\left(\frac{K_{TO}}{J_{TO}}\right)\omega_{m,f}(t)+\left(\frac{k_T}{J_{TO}}\right)\frac{di_{a,f}(t)}{dt}+n(t) \quad (3\text{-}46c)$$

Eq. (3-46c) can be expressed in the sampled vector notation as:

$$\omega_{m,f}^{(2)}(t_k)=\boldsymbol{\mu}_{\text{mech,f}}(t_k)\boldsymbol{\rho}_{\text{mech}}+n(t_k) \qquad (3\text{-}47)$$

where $\boldsymbol{\mu}_{\text{mech,f}}(t_k)$ is a sampled vector of filtered mechanical regressor elements

$$\boldsymbol{\mu}_{\text{mech,f}}(t_k)=\left[-\omega_{m,f}^{(1)}(t_k),-\omega_{m,f}(t_k),i_{a,f}^{(1)}(t_k)\right];$$

$\boldsymbol{\rho}_{\text{mech}}$ is a vector containing the mechanical transfer function parameters:

$$\boldsymbol{\rho}_{\text{mech}}=\left[\left(\frac{B_{TO}}{J_{TO}}\right),\left(\frac{K_{TO}}{J_{TO}}\right),\left(\frac{k_T}{J_{TO}}\right)\right]^T=\left[a_{m,1},a_{m,2},b_{m,11}\right]^T.$$

Therefore, when the SRIVC algorithm is performed on Eq. (3-47), the elements in the mechanical parameter estimation vector, i.e. $\hat{\boldsymbol{\rho}}_{\text{mech}}$, will relate to the mechanical parameters of the DC electro-mechanical turnout systems as follows:

$$\hat{J}_{TO} = \left(\hat{k}_T / \hat{b}_{m,11}\right) = \left(\hat{k}_M / \hat{b}_{m,11}\right), \text{ since } \hat{k}_M = \hat{k}_T$$

$$\hat{B}_{TO} = \left(\hat{a}_{m,1} \times \hat{J}_{TO}\right), \text{ and } \hat{K}_{TO} = \left(\hat{a}_{m,2} \times \hat{J}_{TO}\right). \quad (3\text{-}48)$$

Eqs. (3-35) and (3-48) indicate that such a particular application of the SRIVC algorithm to a permanent magnet DC turnout will provide six physically interpretable parameter estimates, which can be concluded as follows:

- R_A: the electrical resistance of the DC motor.
- L_A: the electrical inductance of the DC motor.
- k_M (which equals k_T for permanent magnet DC motors): the constant value of the DC motor, which relates to its physical characteristics.
- J_{TO}: the total inertia property of the entire DC turnout system.
- B_{TO}: the total damping property of the entire DC turnout system.
- K_{TO}: the total stiffness property of the entire DC turnout system.

Due to their physical interpretations, when the turnout becomes faulty, these parameter estimates should change in correspondence to the associated fault mode (Janot, Young, and Gautier, 2017; Dutta et al., 2020b). The mapping process between the detected parameter changes and associated fault modes is the essential part of fault classification development, which will be developed in Chapters 5 and 6.

In conclusion, this section has demonstrated how the SRIVC algorithm can be employed to estimate the parameters of DC turnout systems. Finally, to illustrate and verify the statistical performance of this algorithm, the last section will compare the parameter estimation performance between the SRIVC and other common algorithms.

3.6 Verification of the SRIVC algorithm performance

The statistical performance of two indirect and two direct parameter estimation algorithms, including LS, SRIV, SVF, and the proposed SRIVC, is evaluated in this section. The data used for this evaluation are generated from the simulation model of a permanent magnet DC turnout system, which was developed in Section 3.4.

In practical parameter estimation of an unknown system, measurement noise is inevitable, while the optimal sampling rate depends on the unknown characteristics of the system of interest as well. Therefore, this comparison analysis implements the Monte-Carlo Simulation (MCS) to assess algorithm performance under two different noise levels (defined by signal-to-noise ratio (SNR) values) and three sampling rates (Gilson, 2015; Wang et al., 2017; Pascu et al., 2019; Pan et al., 2020a; 2020b).

Since the parameters of a simulation model are being estimated, the accuracy of each algorithm can also be assessed as all parameter values used in the simulation are known a priori. To select which algorithm will be investigated further in this thesis, it must provide reliable, consistent, and accurate estimates with a high coefficient of determination, and it must not be highly sensitive to signal noise and sampling rate selection. It should be noted that up to this point, such reliable, consistent, and accurate estimates will only be demonstrated for the simplified linear approximated model.

To begin with, all parameter values used in the simulation are listed in Table 3-1.

Table 3-1 All parameter values used in the simulation model of a permanent magnet DC electro-mechanical turnout system

Model Parameter	Value	Unit
DC motor's armature resistance (R_A)	4.95	Ω
DC motor's armature inductance (L_A)	0.040	H
Voltage constant of the DC motor (k_M)	1.7428	V/(rad/s)

Model Parameter	Value	Unit
Torque constant of the DC motor (k_T)	1.7428	Nm/A
Gear ratio of the gearbox (n)	3.75	-
Point machine's moment of inertia (J_{PM}) where $J_{PM} = J_M + (1/n^2) J_{BS}$	0.1727	Nm/(rad/s)^2
Point machine's coefficient of viscous damping (B_{PM}) where $B_{PM} = B_M + (1/n^2) B_{BS}$	0.0020	Nm/(rad/s)
Lead-pitch distance of the ball-screw (d_{lead})	0.0175	m
Mass of the simulated load (M_{load})	250	kg
Damping coefficient of the simulated load (B_{load})	25,000	N/(m/s)
Stiffness of the simulated load (K_{load})	10,000	N/m

When the simulation begins, a constant input voltage of 110 V is supplied to the DC motor module. The supply does not last for the entire simulation run to allow the declinations of both motor current and speed to be observed. Variables collected during the simulation include the input voltage, $v_s(t_k)$, the armature current, $i_a(t_k)$, the rotational motor shaft speed, $\omega_m(t_k)$, and the load displacement, $x_{load}(t_k)$. The first three are needed for estimation, while the last one is for observation purposes only.

The MCS is applied to conduct a sensitivity analysis on the signal noises and the sampling rate selection. Based on previous test scenarios used in other studies, the SNR levels to be evaluated in this thesis are 15 dB (Soares et al., 2021) and 25 dB (Gilson, 2015), while the three sampling rates are 100 Hz (Du et al., 2019; Romero, Vélez, and Gómez-Mendoza, 2019), 1 kHz (Antong, Dixon, and Ward, 2017), and 25 kHz. The last sampling rate is included based on the capability of the available data acquisition system and the intention to collect near-continuous sampled data.

Based on the parameter values listed in Table 3-1 and the model structures in Eqs. (3-29) and (3-42), the "true" CT electrical and mechanical transfer functions are:

$$\textit{Electrical TF: } i_a(t) = \left(\frac{1}{0.04s + 4.95}\right) v_s(t) + \left(\frac{-1.7428}{0.04s + 4.95}\right) \omega_m(t) \qquad \text{(3-49a)}$$

$$\textit{Mechanical TF: } \omega_m(t) = \left(\frac{1.7428s}{0.1728s^2 + 0.0158s + 0.0055}\right) i_a(t) \qquad \text{(3-49b)}$$

To begin with, all four noise-free variables obtained at 25 kHz are shown in Figure 3-5. These are also compared with measurements collected from a fault-free operation of an HW55S turnout system to justify all parameter values used in this simulation. The input voltage for both scenarios is identical, so only one plot is shown.

Four variables that are observed during the simulation

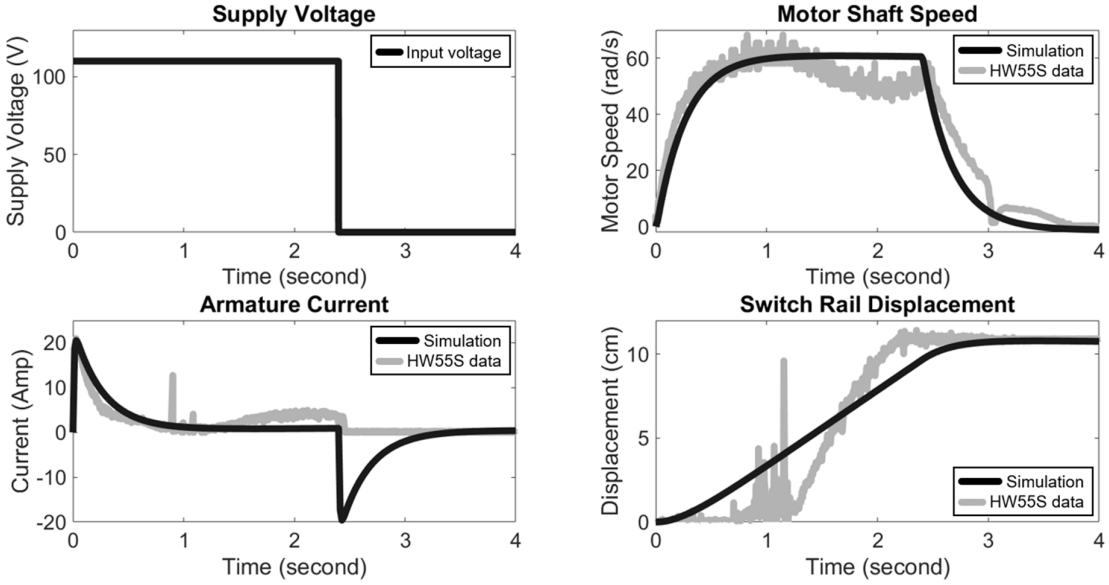


Figure 3-5 Four variables collected from the DC turnout simulation in Figure 3-4, which are also compared to the measurements collected from a fault-free HW55S throw

For brevity, the **accuracy** of each estimation algorithm is evaluated based on the relative differences between the parameter estimates from 5,000 MCS realisations and the values listed in Table 3-1. The **consistency** is graphically illustrated by using boxplots, which can represent the distribution of parameter estimates and their outliers. The **reliability** of each algorithm is assessed from the coefficient of determination (R_T^2)

from every realisation. The threshold of 80%, used in previous studies, is employed to decide which estimation attempt is invalid (Padilla et al., 2019; Dutta et al., 2022).

Since boxplots are often used to represent the parameter estimation results in this thesis, the following figure demonstrates how they do so. Assume there are results from *N* MCS realisations obtained from the same test condition. Once outliers (i.e. results that lie outside the median ±3 SD range) are removed from the samples, the rest can be represented by a boxplot as in Figure 3-6.

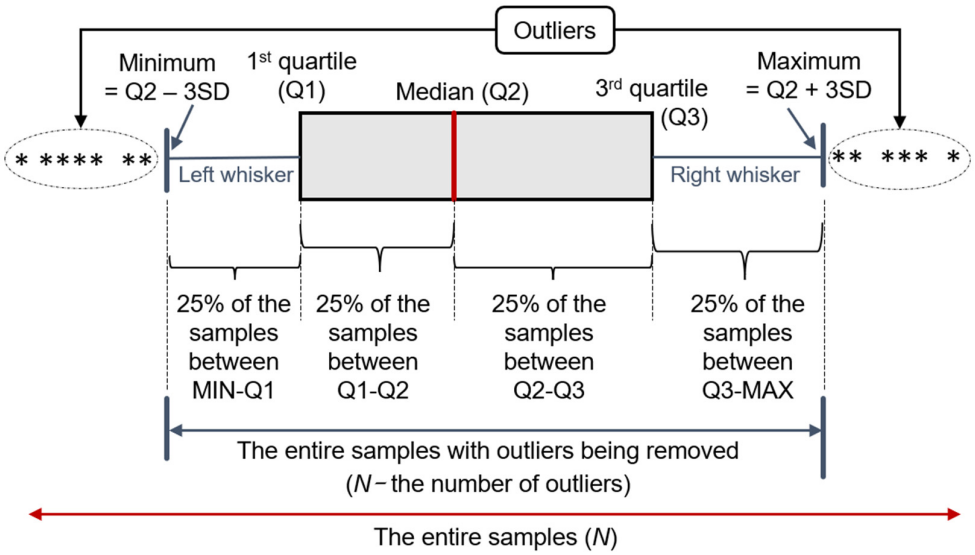
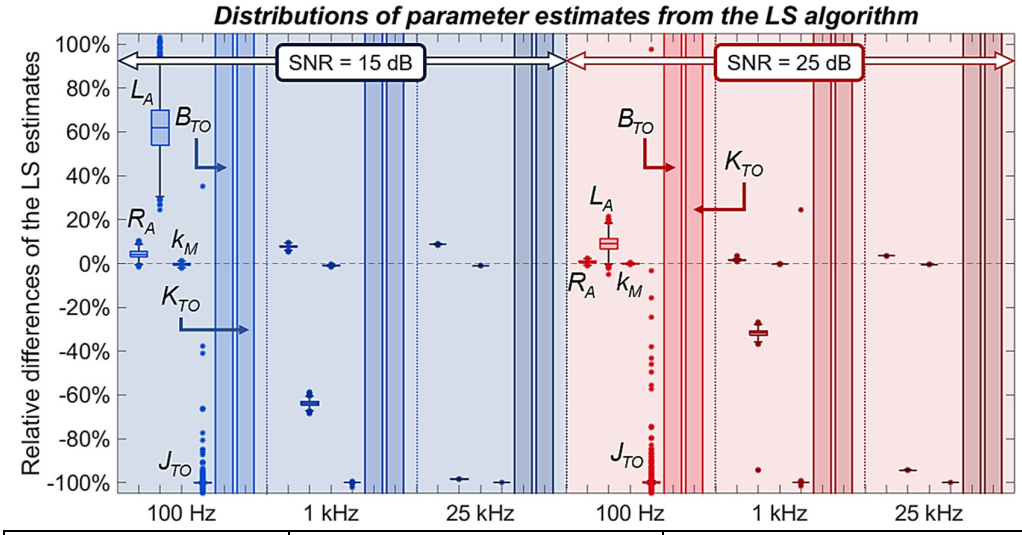


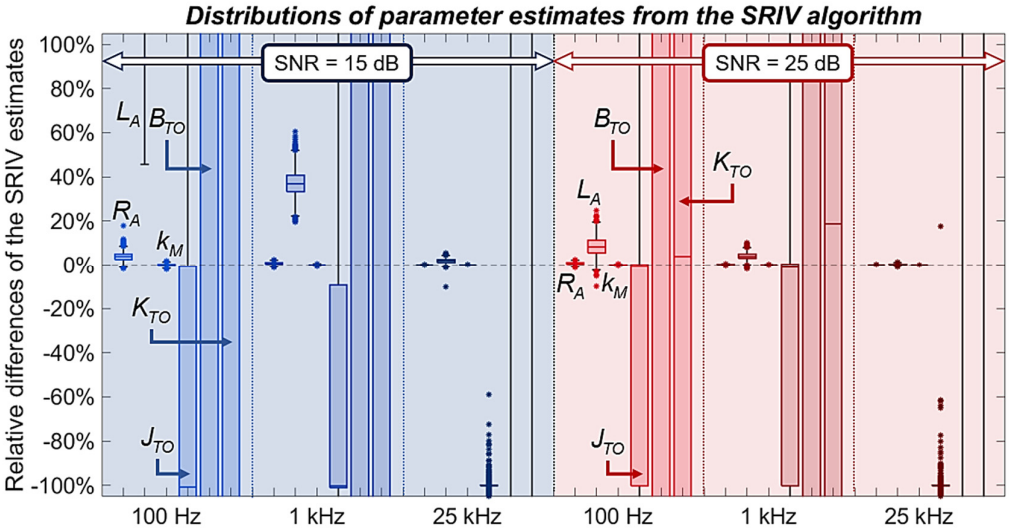
Figure 3-6 Components of a boxplot that will be used to represent a distribution of estimates

The parameter estimation results from LS, SRIV, SVF, and SRIVC algorithms are provided in Figures 3-7 to 3-10, respectively. For each test condition, there are six vertical boxplots representing the estimates in the following order: R_A, L_A, k_M, J_{TO}, B_{TO}, and K_{TO}. Additionally, the accompanying table provides the average R_T^2 from all 5,000 realisations and the number of invalid attempts that achieves R_T^2 less than 80%.



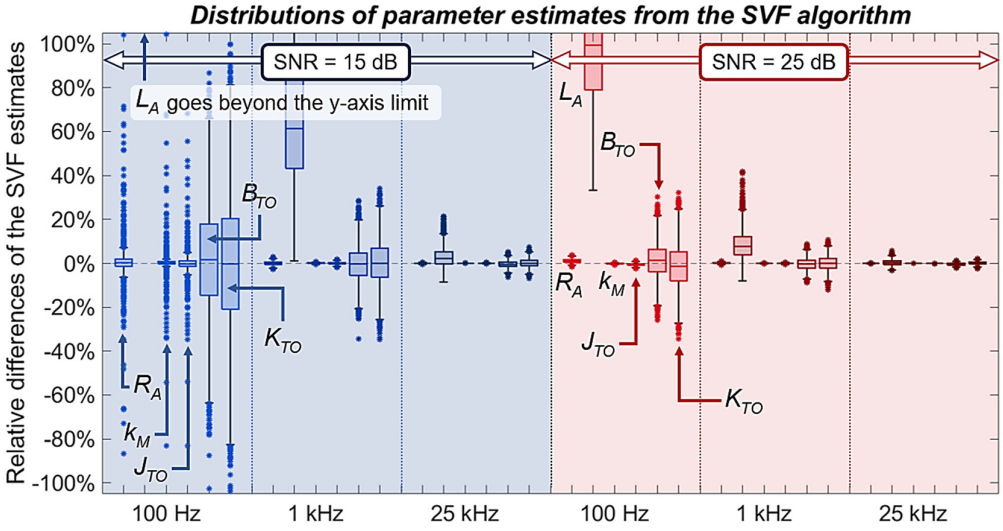
Test Condition	SNR = 15 dB			SNR = 25 dB		
	100 Hz	1 kHz	25 kHz	100 Hz	1 kHz	25 kHz
Results of electrical transfer function estimation						
Average of R_T^2	93.65%	94.15%	92.52%	99.20%	99.29%	97.01%
No. of invalid estimations	0	0	0	0	0	0
Results of mechanical transfer function estimation						
Average of R_T^2	82.54%	15.19%	0%	99.06%	84.59%	3.53%
No. of invalid estimations	760	5,000	5,000	0	500	5,000

Figure 3-7 Parameter estimation results from the LS algorithm



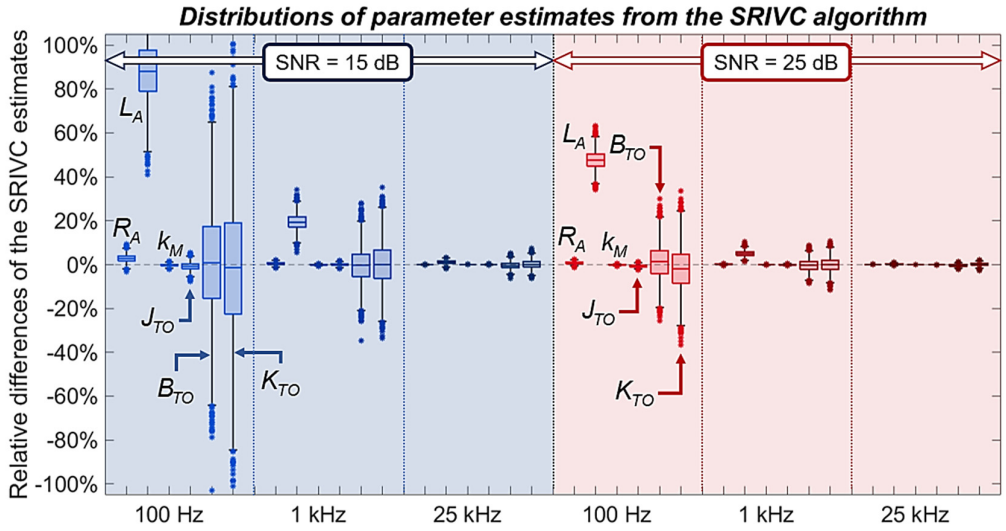
Test Condition	SNR = 15 dB			SNR = 25 dB		
	100 Hz	1 kHz	25 kHz	100 Hz	1 kHz	25 kHz
Results of electrical transfer function estimation						
Average of R_T^2	93.59%	96.30%	96.81%	99.20%	99.52%	99.58%
No. of invalid estimations	0	0	0	0	0	0
Results of mechanical transfer function estimation						
Average of R_T^2	80.43%	80.29%	69.64%	81.18%	77.54%	69.58%
No. of invalid estimations	974	925	1,559	1,223	1,302	1,597

Figure 3-8 Parameter estimation results from the SRIV algorithm



Test Condition	SNR = 15 dB			SNR = 25 dB		
	100 Hz	1 kHz	25 kHz	100 Hz	1 kHz	25 kHz
Results of electrical transfer function estimation						
Average of R_T^2	81.51%	95.90%	96.81%	98.28%	99.51%	99.58%
No. of invalid estimations	1,154	0	0	0	0	0
Results of mechanical transfer function estimation						
Average of R_T^2	96.84%	96.84%	96.84%	99.59%	99.59%	99.59%
No. of invalid estimations	0	0	0	0	0	0

Figure 3-9 Parameter estimation results from the SVF algorithm



Test Condition	SNR = 15 dB			SNR = 25 dB		
	100 Hz	1 kHz	25 kHz	100 Hz	1 kHz	25 kHz
Results of electrical transfer function estimation						
Average of R_T^2	93.61%	96.27%	96.81%	98.73%	99.52%	99.58%
No. of invalid estimations	0	0	0	0	0	0
Results of mechanical transfer function estimation						
Average of R_T^2	96.84%	96.84%	96.84%	99.59%	99.59%	99.59%
No. of invalid estimations	0	0	0	0	0	0

Figure 3-10 Parameter estimation results from the SRIVC algorithm

In Figure 3-7, the most distinguishing features are the estimates of B_{TO} and K_{TO} from the LS algorithm. For every test condition, the distributions of both parameter estimates extend beyond the range of the y-axis, and their median lines are also located outside. These results indicate the inaccuracy and inconsistency of the LS algorithm. Furthermore, the estimates of J_{TO} possess a large relative difference of around −100% for all test conditions. All these properties contribute to a number of invalid estimations of the mechanical transfer function. Meanwhile, although there is no invalid estimation of the electrical transfer function, the asymptotic bias of the L_A estimates are always significant. This poor performance of the LS algorithm leads to the requirement for more sophisticated approaches.

When the SRIV algorithm is employed, estimations of the mechanical transfer function can achieve a higher average R_T^2 and have fewer invalid attempts, especially when measurements are collected with lower sampling rates. Although the data fitting performance is improved, the mechanical parameter estimates are still inaccurate and inconsistent, as shown in Figure 3-8. These undesirable characteristics persist since the resulting CT mechanical transfer function from valid estimations always has one more zero than Eq. (3-49b), making the transformation equations in Eq. (3-48) invalid. Therefore, indirect parameter estimation methodologies are proven not to be suitable for producing physically interpretable parameter estimates for FDD purposes.

Figure 3-9 provides the estimation results from the SVF algorithm when the cut-off frequency is arbitrarily selected as 0.1 Hz (or 0.6283 rad/s). Compared to previous indirect approaches, the mechanical parameter estimates are notably more accurate and more consistent, demonstrating the superiority of direct approaches in estimating physically interpretable parameters. Overall, the estimates possess less bias, fewer

outliers, and are more consistent when the SNR level is higher. This is expected since the normal equation implemented in the SVF algorithm resembles that of the LS.

Interestingly, estimation of the electrical transfer function can achieve a high average R_T^2, but at the 100 Hz 15 dB condition, the invalid attempt rate is about 23%. This is found to be the effect of the cut-off frequency selection since another cut-off frequency of 10 Hz (or 62.83 rad/s) is also investigated, but the results are not shown for brevity. Compared to 0.1 Hz scenarios, there is no longer an invalid electrical transfer function estimation, but it is unfavourable for the estimation of the mechanical transfer function instead. The comparison results highlight the requirement of system knowledge to select an appropriate cut-off frequency when using the SVF algorithm.

Finally, results from the proposed SRIVC algorithm are given in Figure 3-10. It provides the most accurate and consistent parameter estimates for both electrical and mechanical transfer functions with no invalid attempt. The performance at the 25 kHz conditions is also the most effective one, as expected for direct parameter estimation. Compared to the SVF algorithm, there are noticeable improvements in terms of both asymptotic bias and consistency, while it does not require any manual setting. Due to its automated nature and excellent estimation performance, the practical application of the SRIVC algorithm for railway turnout FDD is investigated throughout this thesis.

3.7 Conclusions

This chapter introduces the parameter estimation principle that provides FDD capabilities based on the change detection of parameter estimates. Two approaches (i.e. indirect and direct) that are distinguished by the type of model being estimated were elaborated. Due to the effects of sampling rate on the estimates and the model domain conversion required in indirect approaches, direct parameter estimation was

expected to be more practical for railway turnout FDD. Two of its common problems, namely asymptotic bias and the requirement of filtering, were also pointed out. Based on previous successful applications of the SRIVC algorithm, especially in handling these two issues, it is proposed to be implemented in this thesis.

To evaluate its effectiveness, a simulation model of a DC turnout system was formulated, and the application of the SRIVC algorithm to the model was described. Finally, its statistical performance was compared to other algorithms, including LS, SRIV, and SVF. An MCS was employed to analyse their sensitivity to measurement noises and the sampling rate used as well. The evaluation concluded that the SRIVC algorithm is capable of providing the most reliable, accurate, and consistent estimation results while being less susceptible to noise and sampling rate selection.

In conclusion, continuing from Chapter 2, this chapter particularly chooses the SRIVC parameter estimation algorithm to be implemented throughout this thesis. The selection is based on the performance comparison between the SRIVC and other commonly used parameter estimation algorithms. With a set of assumptions and data obtained from a turnout simulation model, the statistical performance of the SRIVC algorithm is verified to be superior to the others. The following chapter continues the investigation of the SRIVC application for railway turnout FDD when the turnout is represented by a more realistic model. The parameter change detection capability will also be evaluated before the methodology is applied to a working turnout system.

CHAPTER 4 VALIDATION OF THE PROPOSED TRANSFER FUNCTION MODELS TO REPRESENT SWITCH RAIL DYNAMICS

The turnout system dynamics are considered in more detail in this chapter. First, high-fidelity models that have been previously used to represent them are introduced. Then, the load's transfer function proposed in Chapter 3, also shown in Figure 4-1, is validated. This is to assess whether this transfer function can still faithfully represent the switch rail's bending dynamics. Finally, the capability of the SRIVC algorithm to detect parameter changes, applied to the validated transfer functions, is evaluated as it is the primary objective of parameter estimation applications for FDD purposes.

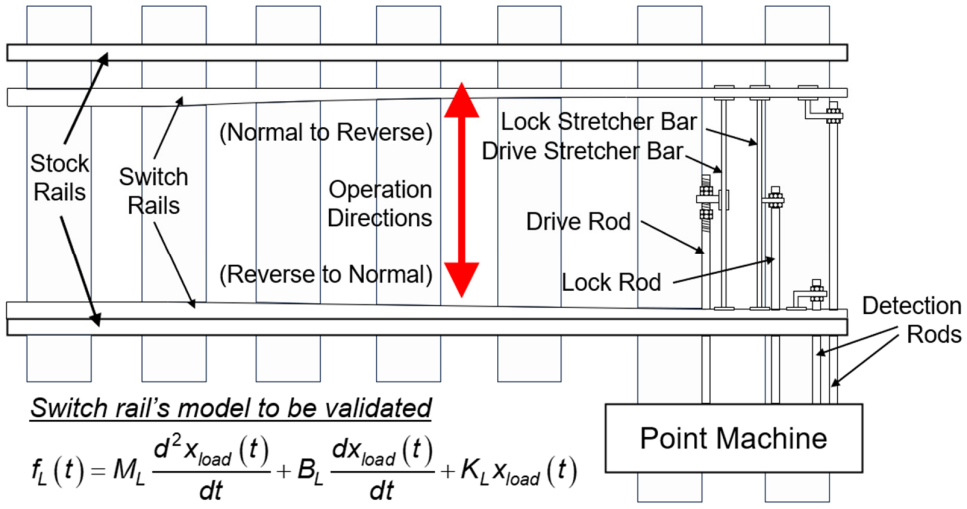


Figure 4-1 The switch rail assembly that the point machine can only actuate at the drive stretcher bar and the switch rail's differential equation to be validated in this chapter

4.1 Introduction to switch rail bending dynamics

Switch rails are movable track elements that have cross-sections identical to those of stock rails at the heel end, which become narrower towards the front toe. They

are brought together by stretcher bars, connectors, and other fasteners to maintain the appropriate separation between them (i.e. the track gauge), as shown in Figure 4-2.

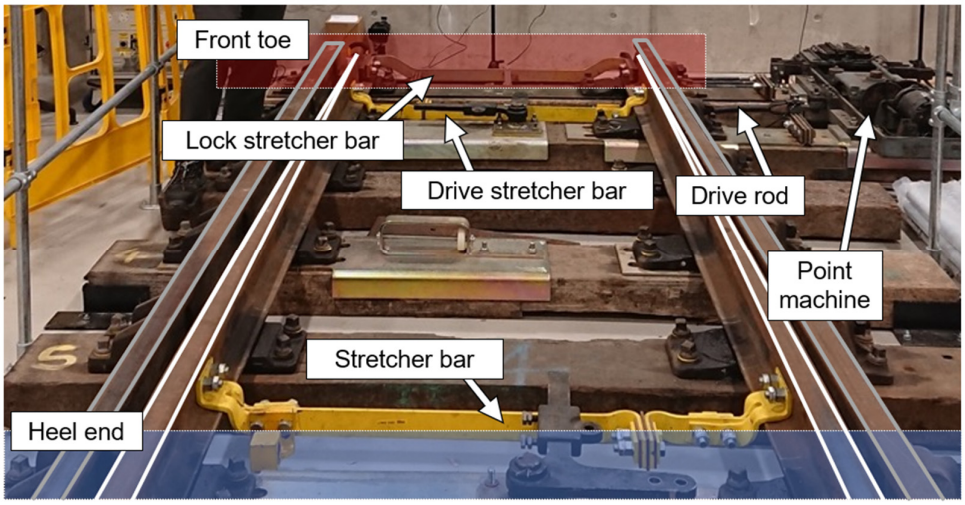


Figure 4-2 The changing geometry of the switch rails that start from cross-sections identical to those of the stock rails at the heel end and become narrower towards the front toe

When the route setting requires the switch rails to be re-positioned, a command will be sent to operate the turnout. It will then actuate the switch rails to the other end position and lock them in place. Since the actuating force applied to the switch rails is only concentrated at the drive stretcher bar, it will bend the rails instead of moving them linearly. Furthermore, a switch rail is essentially a steel beam with non-uniform cross-sections that possess varying properties along its length. It is relatively stiff at the heel end due to its larger cross-sectional area while being much more flexible at the front toe. Therefore, the switch rail's actuation is expected to be adequately represented by the bending dynamics of a non-uniform cantilever beam when a concentrated force is given at one particular position (Roberts et al., 2002; Bemment et al., 2017).

Previous studies have attempted to create mathematical models to understand and represent such dynamic behaviours of switch rails (Oyebande and Renfrew, 2007; Bai, 2010; Matei et al., 2015; Sarmiento-Carnevali et al., 2017; Olaby et al., 2021). However, the articles of Dutta et al. (2018), Li et al. (2022), and other related works

are the only ones that have shown successful validations with measurements collected from a working HPSS turnout provided by Network Rail, UK.

To begin with, Dutta et al. (2018) modelled the switch rail's bending dynamics based on the Finite Element Analysis (FEA) principle. The material was isotropic steel, and the models were constructed from second-order quadratic tetrahedral elements. Once these FEA models were converted into flexible bodies in Abaqus software, they were used in SIMPACK to form multibody simulations. Finally, the SIMAT interface was used to create a co-simulation platform that enabled data exchange between the SIMPACK and Simulink® models during the simulation, as shown in Figure 4-3.

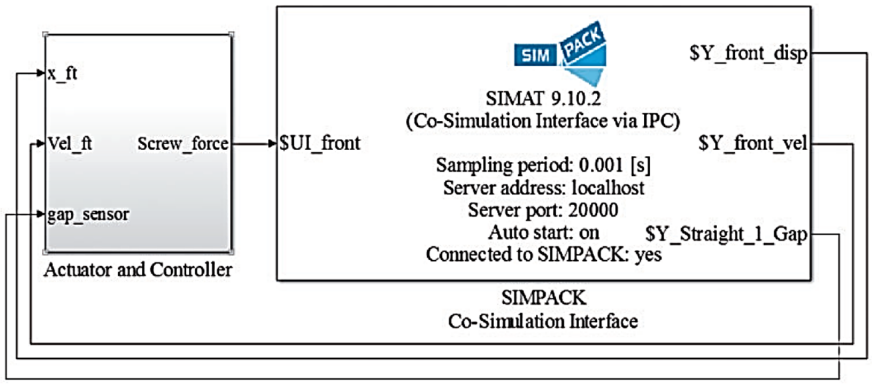


Figure 4-3 The co-simulation environment between SIMPACK and Simulink® using the SIMAT interface to allow data exchange between the two models (Dutta et al., 2020a)

This co-simulation model could represent the switch rail's bending dynamics to such an extent that it could facilitate the controller design for an HPSS self-adjusting turnout (Dutta et al., 2020a) and for a "REPOINT-Light" turnout system (Olaby et al., 2021). However, it required a lot of computational resources to simulate. As a result, Li et al. (2022) developed another model within a single simulation environment that can be simulated on a near real-time basis instead. Since this model is more intuitive, it will be briefly explained to introduce the modelling process of the switch rails.

First, the size and shape of the cross sections along the switch rail's length are not consistent, which can be illustrated in Figure 4-4.

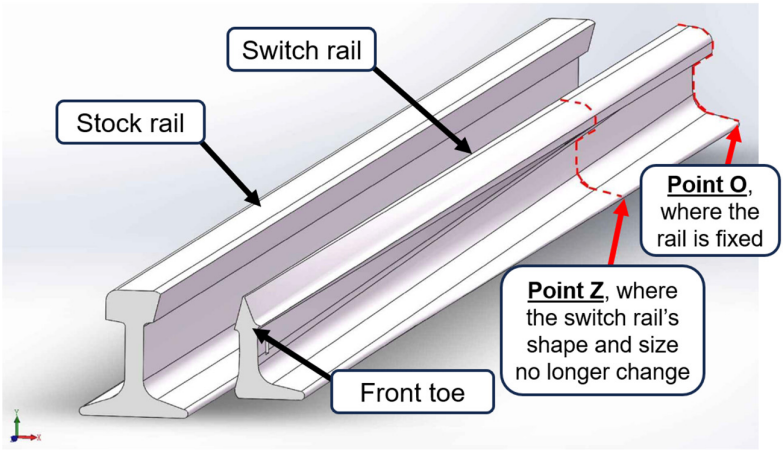


Figure 4-4 A 3D model of the stock rail and the switch rail demonstrating the changing of size and shape of cross sections from Point Z to the front toe (Li et al., 2022)

Once the switch rail is divided into multiple discrete FEA elements, their local stiffness and mass matrices for each element are as follows (Kwon and Bang, 2000):

$$[k]_i = \frac{EI_i}{L_e^3}\begin{bmatrix} 12 & 6L_e & -12 & 6L_e \\ 6L_e & 4L_e^2 & -6L_e & 2L_e^2 \\ -12 & -6L_e & 12 & -6L_e \\ 6L_e & 2L_e^2 & -6L_e & 4L_e^2 \end{bmatrix} \tag{4-1}$$

$$[m]_i = \frac{\rho A_i L_e}{420}\begin{bmatrix} 156 & 22L_e & 54 & -13L_e \\ 22L_e & 4L_e^2 & 13L_e & -3L_e^2 \\ 54 & 13L_e & 156 & -22L_e \\ -13L_e & -3L_e^2 & -22L_e & 4L_e^2 \end{bmatrix} \tag{4-2}$$

where E is the Young's modulus constant of the isotropic steel = 2×10^{11} N/m^2;

ρ is the mass density of the isotropic steel = 7,850 kg/m^3;

L_e is the length of each FEA element (a total of 20 elements) = 0.4605 m;

L is the total movable length of the switch rail to be modelled = 9.21 m;

I_i is the area moment of inertia of the i^{th} element that varies with the position:

$$I_i = \begin{cases} I_0\left[1-\left(\frac{1}{4}\right)^2\left(\frac{L_i - Z}{L - Z}\right)\right] & ;(L > L_i > Z) \\ I_0 & ;(L_i \le Z) \end{cases} \tag{4-3}$$

I_0 is the area moment of inertia at the heel end = 5.6×10^{-6} m^4;

L_i is the reference position of the i^{th} element (referred from Point O);

Z is the position where the size and shape of the cross sections start to change (measured from Point O) = 5.037 m;

A_i is the cross-sectional area of the i^{th} element that varies with the position:

$$A_i = \begin{cases} A_0\left[1-\left(\frac{1}{4}\right)\left(\frac{L_i - Z}{L - Z}\right)\right] & ;(L > L_i > Z) \\ A_0 & ;(L_i \le Z) \end{cases} \quad (4\text{-}4)$$

A_0 is the cross-sectional area at the heel end = 9.334×10^{-3} m^2.

Then, the global matrices for the entire switch rail can be formulated from their associated local matrices (Dutta et al., 2023). To include the damping effects in the FEA model as well, proportional Rayleigh's damping was applied in their development as follows (Ling and Haldar, 2004; Panzer et al., 2009):

$$[C] = a_0[M] + a_1[K] \quad (4\text{-}5)$$

where $[M]$ is the global mass matrix of the FEA switch rail model;

$[C]$ is the global damping matrix of the FEA switch rail model;

$[K]$ is the global stiffness matrix of the FEA switch rail model;

a_0 and a_1 are Rayleigh's coefficients that were obtained via experiments (Miller et al., 2017). According to Li et al. (2023), $a_0 = 1$ and $a_1 = 0.6$.

The FEA model of one switch rail has now been formulated. However, there are two switch rails coupled to each other by stretcher bars, where the actuating force will be applied to; see Figure 4-5. So, the differential equations governing the dynamics of the element that is being actuated on both switch rails can be expressed as follows:

For mass m_1, which represents the connected element of the first switch rail:

$$u - (k_1 + k_2)x_1 - (b_1 + b_2)\dot{x}_1 + k_2x_2 + b_2\dot{x}_2 = m_1\ddot{x}_1 \quad (4\text{-}6a)$$

For mass m_2, which represents the connected element of the second switch rail:

$$-(k_2 + k_3)x_2 - (b_2 + b_3)\dot{x}_2 + k_2x_1 + b_2\dot{x}_1 = m_2\ddot{x}_2 \quad (4\text{-}6b)$$

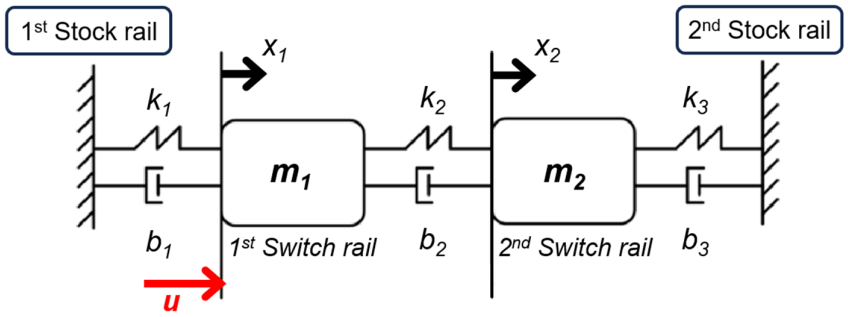


Figure 4-5 Schematic of the connections between two switch rails that are connected by a drive stretcher bar and being confined between two stock rails (Li et al., 2022)

In the figure, the couplings at 1 (i.e. k_1 and b_1) and 3 (i.e. k_3 and b_3) represent the contact between the closing stock rail and the switch rail at the beginning and the end of each throw. The coupling at 2 (i.e. k_2 and b_2) represents the drive stretcher bar that is connected to both switch rails and is the position where the actuating force u will be provided from the point machine. Therefore, it should be ideally stiff (i.e. $k_2 \to \infty$ and $b_2 \to \infty$) when the turnout system is in fault-free conditions.

Finally, as multiple FEA elements are all connected along the rail, an effect on one element will influence all others as well. The differential equations describing the dynamics of the entire switch rail assembly can be expressed as follows:

$$[M]\ddot{\vec{X}} + [C]\dot{\vec{X}} + [K]\vec{X} = \vec{F} - \vec{f}_R \tag{4-7}$$

where $\vec{X}$ is a vector containing the displacement of each element;

$\dot{\vec{X}}$ and $\ddot{\vec{X}}$ are vectors containing the velocity and acceleration of each element;

$\vec{F}$ is a vector consisting of the actuating force applied to each element;

$\vec{f}_R$ is a vector consisting of horizontal friction forces that resist the movement of each element: for every odd element, $f_{R,i} = \mu M_{rail} g \times \left[A_i / sum(A_i)\right]$;

μ is the coefficient of friction between the switch rail and the slide chairs;

M_{rail} is the total mass of the switch rail that is being modelled = 276 kg;

g is the gravitational acceleration = 9.81 kg·m/s^2.

As previously mentioned, both Dutta's co-simulation and Li's FEA model have already been validated with measurements obtained from a fault-free working HPSS turnout (Dutta et al., 2018; Li et al., 2022). The validation concluded that although Li's FEA model was developed with a simpler approach, it could still faithfully represent the switch rail's bending dynamics. Thus, it will now replace the mass-spring-damper load model previously used in Chapter 3 for more realistic simulations. Its first application will be to validate the representability of both the turnout's electrical and mechanical transfer functions formulated in Chapter 3, i.e. Eqs. (3-29) and (3-42).

4.2 Validation of the turnout's transfer function models

The success of the SRIVC algorithm demonstrated in Chapter 3 was primarily because a mass-spring-damper assembly was used to represent the switch rails. This simplification allowed the model structures to be known deterministically and identical to those of the system to be estimated, resulting in the following transfer functions:

$$\textit{Electrical TF: } I_a(s) = \left(\frac{1}{sL_A + R_A}\right) V_s(s) + \left(\frac{-k_M}{sL_A + R_A}\right) \Omega_m(s) \tag{3-29}$$

$$\textit{Mechanical TF: } \Omega_m(s) = \left[\frac{k_T}{s^2 J_{TO} + sB_{TO} + K_{TO}}\right] sI_a(s) \tag{3-42}$$

In this section, the validity of both transfer functions is evaluated when Li's FEA model replaces the mass-spring-damper load. From Eqs. (4-1) to (4-7), the integration to the previous DC turnout simulation model in Simulink® is illustrated in Figure 4-6.

The top section represents the first switch rail, while the bottom represents the second one. The middle part corresponds to the couplings at all three positions (refer to Figure 4-5). The inputs are the actuating force and friction force vectors, i.e. $\vec{F}$ and $\vec{f}_R$, while the outputs are displacement and velocity of the switch rail's front toe.

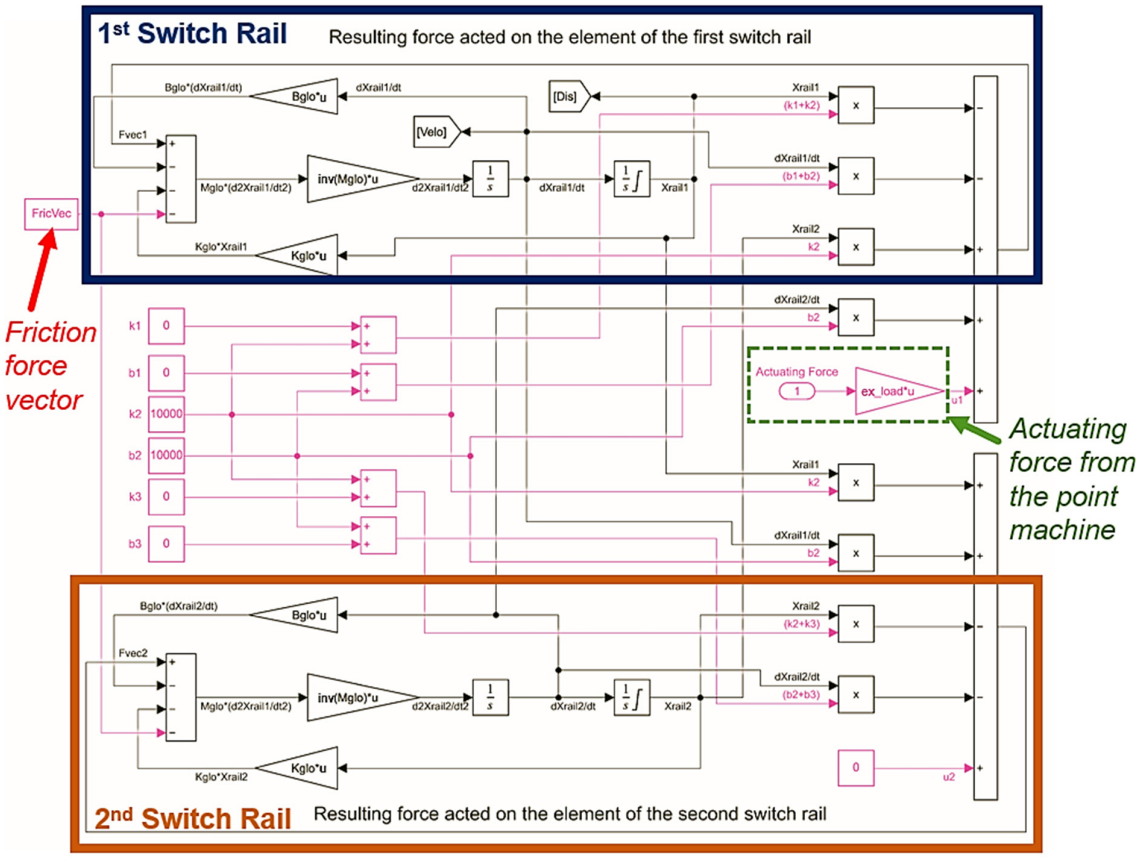


Figure 4-6 Li's FEA model that is integrated into the DC turnout simulation model in order to represent the switch rail's bending dynamics

For each simulation run, the same three variables, i.e. the input voltage, motor current, and motor speed, are collected to estimate the parameters of Eqs. (3-29) and (3-42) using the SRIVC algorithm. In this circumstance, the integration of Li's FEA model will cause the linearised transfer functions of the turnout simulation model to have different structures compared to those of Eqs. (3-29) and (3-42). Therefore, the "true" transfer functions will be unattainable, resulting in the accuracy of the parameter estimates being incomputable as well, whereas the coefficient of determination will be the only viable criterion. Using the same parameter values for simulation as in Li et al. (2023), the SRIVC results are provided in Table 4-1 and Figure 4-7.

Table 4-1 SRIVC estimated parameters of Eqs. (3-29) and (3-42) when Li's FEA model is used to represent the switch rail's dynamics instead of a mass-spring-damper assembly

Electrical Parameter	***Estimation result***	***Mechanical Parameter***	***Estimation result***
R_A – Resistance (Ω)	5.0	J_{TO} – Turnout's Inertia (Nm/(rad/s^2))	10.6×10^{-3}
L_A – Inductance (mH)	10.80	B_{TO} – Turnout's Damping (Nm/(rad/s))	1.4×10^{-3}
k_M – Constants (V/(rad/s))	0.2865	K_{TO} – Turnout's Stiffness (Nm/rad)	2.3×10^{-3}

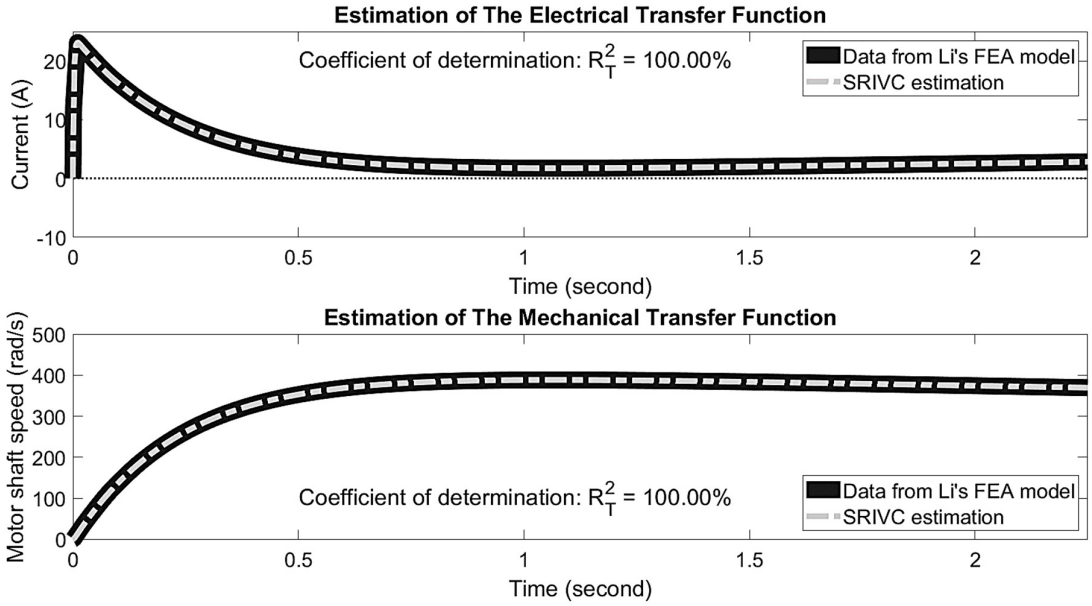


Figure 4-7 Comparison between the simulation data that integrated Li's FEA model and the SRIVC estimation results, with the coefficient of determination

Figure 4-7 shows that the responses of the proposed transfer functions with the SRIVC parameter estimates behave identically to Li's FEA model, achieving a unity R_T^2. Additionally, every system parameter estimate in Table 4-1 is within its reasonable range, making it still valid and physically meaningful as well. Hence, these transfer functions are successfully validated and can be used to govern the turnout dynamics.

In the following section, the second validation will be conducted, where it will observe if the system parameter estimates change correspondingly to the changes in parameters made during fault simulations. Validating such parameter change detection capability will be instrumental for the later stages of the FDD algorithm development.

4.3 Parameter change detection of the SRIVC algorithm

In this section, multiple fault modes, including ones that are difficult to simulate on a real turnout system, are simulated by changing the associated parameter values used in the simulation. For every test scenario, the input voltage, motor current, and motor speed are collected for SRIVC parameter estimation. The results are evaluated to see whether the algorithm can detect changes in the corresponding parameters.

Based on the validation process conducted in the previous section, the system parameter estimates of the fault-free condition have already been extracted and are listed in Table 4-1. Therefore, the estimation results obtained from each fault mode are compared to this list to calculate the relative differences for evaluation. Referring to the related works of Dixon (2000), Hamadache et al. (2019b), and Dutta et al. (2022), the fault modes considered in this validation are listed as follows:

- Changes in both armature resistance and inductance to simulate motor winding-related faults, which affect multiple parameters at the same time;
- Changes in the mechanical parameters of the point machine to simulate internal wear and tear faults in its driving mechanism;
- Changes in the motor constant to simulate physically related motor faults;
- Changes in the coefficient of friction to simulate improper sliding surface faults.

To assess the performance of parameter change detection capability, an MCS is employed in this evaluation as well. That is, for each test condition, there are 5,000 realisations. The measurements are sampled at 25 kHz and contaminated by signal noises with an SNR of 15 dB. The assessment is then focused on the distributions of these parameter estimation results.

4.3.1 Motor winding fault simulation

In this case, both armature resistance and inductance are varied, which could be the effects of incipient inter-turn short circuits or varying motor temperatures. Based on the fault-free values (5 Ω and 10.8 mH), the degrees of change in both parameters that are imposed to conduct fault simulation are 1.5 times, 2 times, 75%, and 50%. The SRIVC estimation results of this fault are illustrated in Figure 4-8.

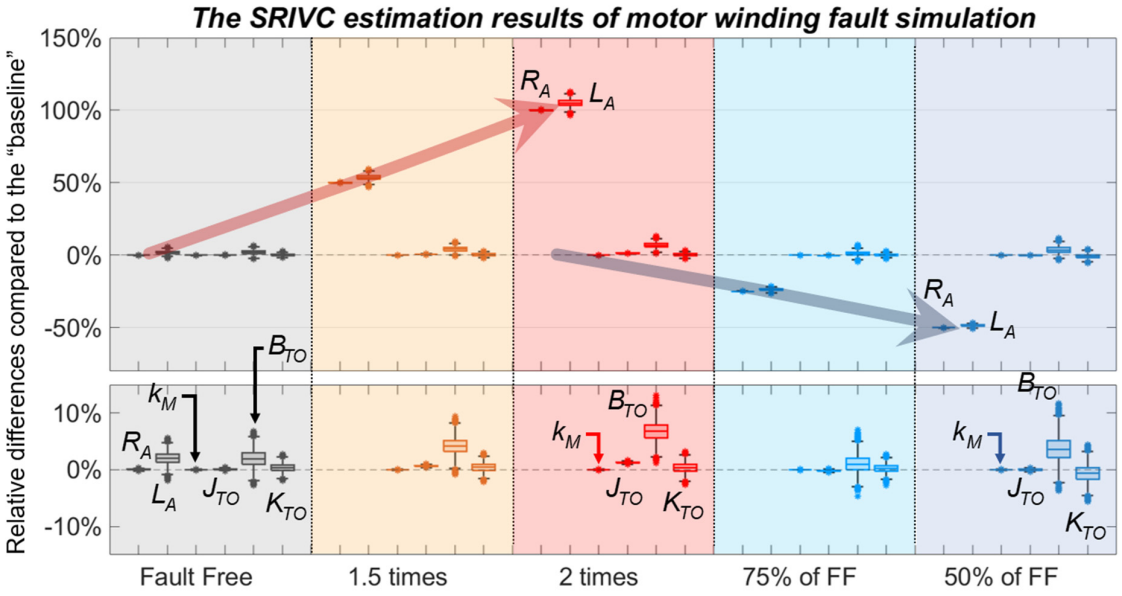


Figure 4-8 SRIVC estimation results of the measurements collected from the simulated motor winding faults by changing the armature resistance and inductance

The presentation format in this section is similar to the algorithm comparison in Chapter 3, except for the additional trend lines. That is, each test condition contains six boxplots representing the estimates of R_A, L_A, k_M, J_{TO}, B_{TO}, and K_{TO}, respectively. What is more, there are two panels with different y-axis ranges. The top one has a wider range to cover the changes in the corresponding system parameters, whereas the bottom one indicates which parameters are not affected by the simulated faults.

For motor winding fault simulations, the top panel clearly demonstrates that the estimates of both R_A and L_A correlate with this fault. The bottom panel also confirms

that other unrelated parameters do not divert from their fault-free “baseline” values. Although the estimates of B_{TO} are the most inconsistent, their distribution only covers from around +10% to −5%, which is negligible compared to the changes in R_A and L_A.

4.3.2 Wear and tear in the point machine fault simulation

The previous fault simulation only applied changes to the electrical parameters. In this case, the inertia and damping coefficients of the motor inside the point machine are increased to four different levels to simulate its incipient deterioration. Based on their fault-free values, they are multiplied by factors of 1.25, 1.5, 1.75, and 2. The SRIVC estimation results of these modifications are displayed in Figure 4-9.

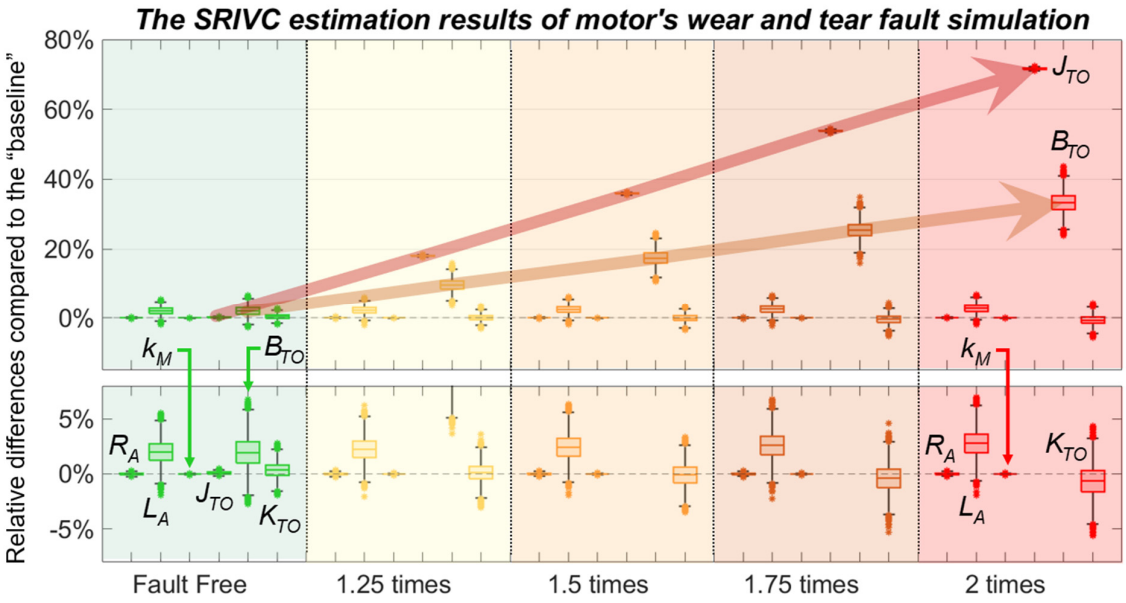


Figure 4-9 SRIVC estimation results of the measurements collected from the simulated motor’s wear and tear faults by changing the motor’s inertia and damping coefficients

Once again, the bottom panel clearly shows that all other unrelated parameters are at the same values as their “baseline”. Notable changes only appear in J_{TO} and B_{TO}. Since both of them equal the combination of that property contributed by the motor, the ball screw, and the switch rails, changing the motor’s parameters will affect that property of the entire turnout system. Although the degrees of increase in both J_M

and B_M are identical, the top panel indicates that the motor contributes the inertia effect in a higher proportion compared to the damping. Therefore, the SRIVC estimates of J_{TO} demonstrate a faster-increasing rate than that of B_{TO}.

4.3.3 Physically-related motor fault simulation

This simulation duplicates a major structural fault inside the DC motor, which is difficult to simulate in a real turnout system and may permanently damage it. As shown in Table 4-1, the fault-free motor constant is 0.2865 V/(rad/s). It is changed into four different levels to simulate this fault mode, namely, 1.5 times, 2 times, 75%, and 50%. The SRIVC estimation results of these fault simulations are shown in Figure 4-10.

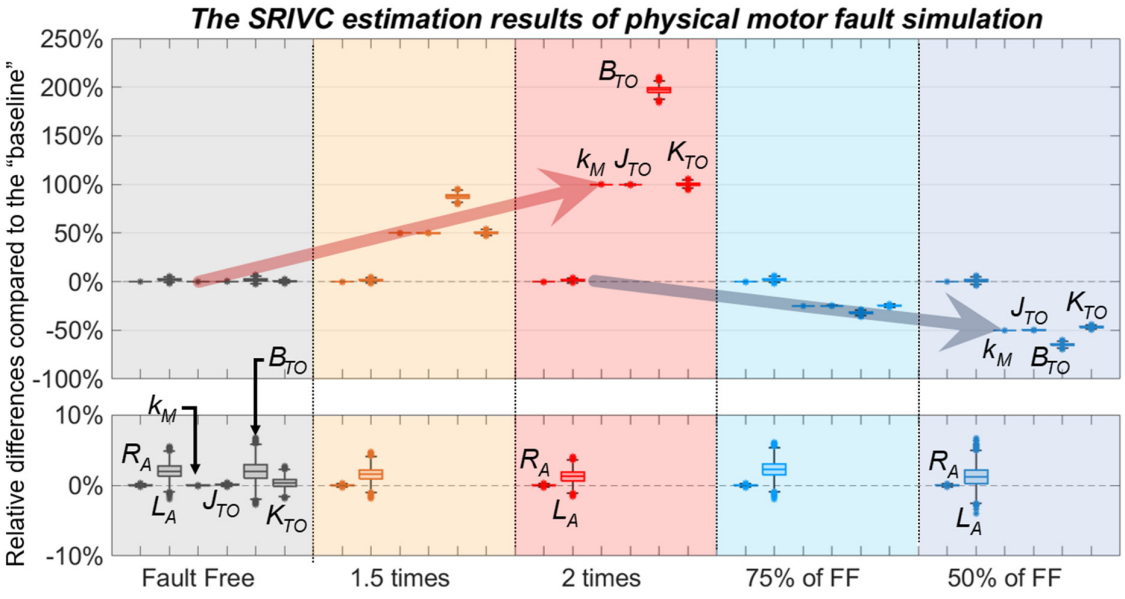


Figure 4-10 SRIVC estimation results of the measurements collected from the simulated physically related motor faults by changing the motor's voltage constant

In this scenario, R_A and L_A are not affected by this fault simulation. Instead, changes in the motor constant will influence the estimates of itself and all mechanical parameters. This may seem counterintuitive at first, so it requires further evaluation.

To begin with, a permanent magnet DC motor is modelled in this simulation, so it possesses the relationship of $k_M = k_T$. Then, Eq. (3-48) is repeated here as follows:

$$\hat{J}_{TO} = \left(\hat{k}_T / \hat{b}_{m,11}\right) = \left(\hat{k}_M / \hat{b}_{m,11}\right),\ \hat{B}_{TO} = \left(\hat{a}_{m,1} \times \hat{J}_{TO}\right),\ \hat{K}_{TO} = \left(\hat{a}_{m,2} \times \hat{J}_{TO}\right) \quad (3\text{-}48)$$

These are transformation equations that transform coefficients of the transfer function into the turnout's parameters. Therefore, changes in k_M will affect the estimate of itself and k_T, and ultimately influence all mechanical parameter estimates as well.

From the boxplots in the top panel, it can be concluded that when the estimates of k_M are changing in the same order of magnitude as all mechanical parameters, it is highly probable that faults that affect the motor constant have occurred.

4.3.4 Improper sliding surface fault simulation

The sliding surface that is normally used to ease the switch rail's actuation may consist of either lubricated slide chairs or rollers. To simulate the situations where the sliding surface is improper, e.g. contaminated, damaged, or not lubricated, this fault simulation focuses on the coefficient of friction. Initially, in the fault-free condition, it is set to equal 0.05. Here, it is increased to four different levels: 0.1, 0.2, 0.4, and 0.8. The SRIVC estimation results from these changes are illustrated in Figure 4-11.

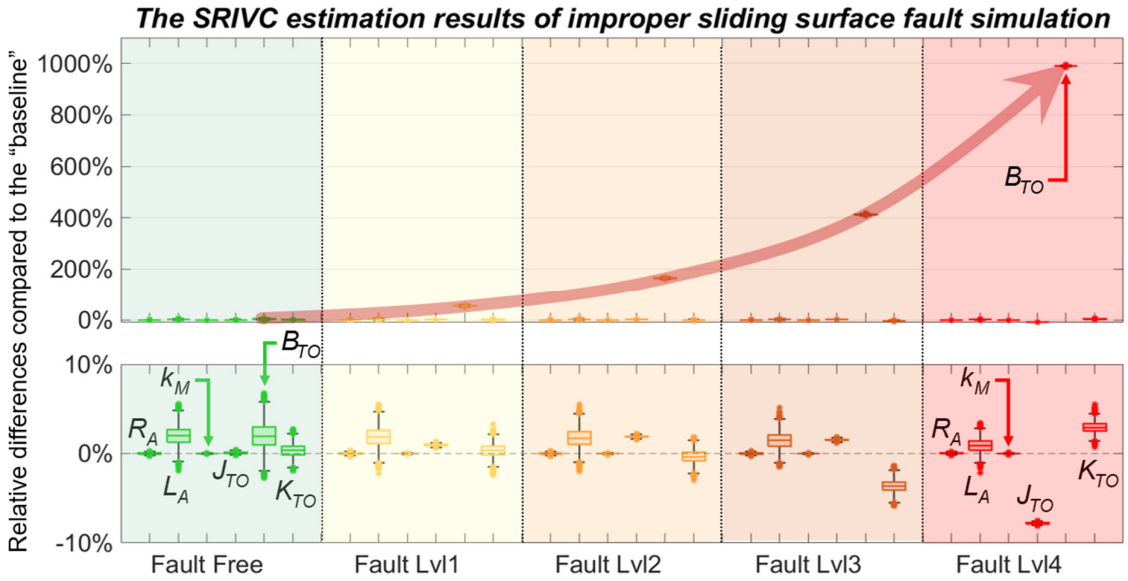


Figure 4-11 SRIVC estimation results of the measurements collected from the simulated improper sliding surface faults by changing the coefficient of friction

The bottom panel of the figure clearly illustrates that all system parameters are relatively unchanged, except the turnout's damping (B_{TO}). Because the coefficient of friction was increased from one level to another to simulate higher severities, B_{TO} also increased exponentially, showing strongly distinguished correlations between the two.

4.4 Conclusions

The successful application of the SRIVC algorithm in Chapter 3 was due to the fact that the structure of transfer function models could be formulated to be identical to those of the simulation model that is being estimated. It was feasible since a mass-spring-damper assembly was used to represent the switch rail's bending dynamics.

This chapter reduces the gap between a simulation and a real turnout system by investigating the turnout dynamics in more detail. First, two high-fidelity models previously validated and used in other applications were introduced, one of which was developed from a more intuitive approach. So, it was used to replace the mass-spring-damper load in the simulation to represent the switch rail's dynamics more realistically.

Then, the same transfer functions proposed in Chapter 3, which were proven to be effective for SRIVC parameter estimation, were validated to see whether they could still be used to represent the turnout dynamics faithfully. This started by obtaining data when the simulation model was run in fault-free conditions. The SRIVC algorithm was then implemented, which resulted in a unity coefficient of determination, indicating an excellent data fitting performance. Furthermore, the turnout parameter estimates were all positive and in their reasonable ranges, making them physically interpretable. Thus, the representability of both transfer functions was successfully validated to be adequate in describing the turnout system dynamics.

The second validation focused on the parameter change detection capability of the methodology. Multiple fault modes were simulated by changing the parameter values used in the simulation. Once the SRIVC parameter estimation was conducted on measurements obtained from each test condition, the estimates were examined to see whether they could reflect the changes. The results showed that the combination of the validated transfer functions and the SRIVC parameter estimation algorithm can detect changes in the corresponding parameters effectively.

In summary, this chapter first modifies the simulation model to be more realistic by integrating the non-linear switch rail model. Then, the same linear transfer functions and the SRIVC algorithm verified in Chapter 3 are applied to the simulation data. With an excellent data-fitting performance, this methodology is validated. After that, faults are injected into the simulation model and the SRIVC algorithm is used to estimate the affected parameters. The results validate that the methodology can also accurately detect parameter changes. Due to such successful validations, in the following chapter, this methodology is going to be applied to the measurements collected from a real HW55S turnout system available at the University of Birmingham. This is to assess the preliminary practicality of the parameter estimation-based FDD methodology, which has never been carried out on a working turnout system in any research before.

CHAPTER 5 PARAMETER CHANGE DETECTION CONDUCTED ON AN HW55S TURNOUT SYSTEM WITH SIMULATED FAULTS

Successful application of the validated transfer functions and SRIVC parameter estimation to turnout simulation models have been demonstrated in Chapter 4. Since most algorithms usually work well with simulated data but may fail to do so with data obtained from a real system (Åström and Eykhoff, 1971), in this chapter, the proposed methodology's performance when it is applied to a real HW55S turnout system will be validated. One of the HW55S turnouts is available in the laboratory at the University of Birmingham; besides, many of its descendants (i.e. HW1000 and HW2000 series) are still in use in many UK railways today. Examples are displayed in Figure 5-1.



Figure 5-1 Examples of HW turnout systems currently used in UK railways (Author, 2023)

This process presents the novelty of this thesis since no study has ever applied the SRIVC parameter estimation algorithm to a working turnout system before. It also aims to validate if the parameter estimation is adequately flexible with respect to the assumptions used in previous chapters. These include known model structures of the system to be estimated and the statistical characteristics of measurement noises.

First, the HW55S turnout system and the experimental setup are introduced. Then, the SRIVC algorithm will be applied to the measurements collected from it while operating in fault-free conditions to create the “baseline” for each parameter. Following this, five fault modes: overdrive, obstruction, underdrive, abnormal input voltage, and improper sliding surface faults, as previously described in Chapter 2, will be simulated on this turnout. The objective is to compare the turnout parameter estimates between the baseline and those acquired from each fault simulation.

At the end of this chapter, a knowledge-base table indicating the relationships between the fault modes and the associated turnout parameters will be created, which will be fed into the FDD algorithm development in Chapter 6.

5.1 HW55S turnout system and experimental setup

The available HW55S turnout system is illustrated in Figure 5-2, which has the arrangement of the left hand normally closed and is installed in a non-ballasted track. At the normally closed position, if a train is travelling from the facing direction, it will continue on the “normal” route. Otherwise, it will divert to the “reverse” route instead.

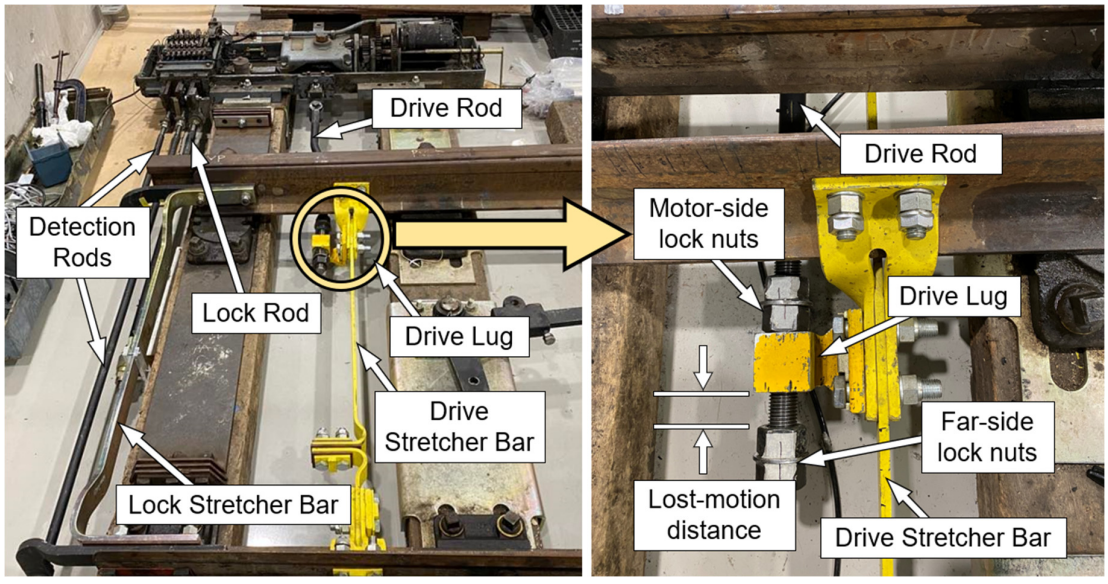


Figure 5-2 Available HW55S turnout system and its connections to the switch rail assembly

The HW55S driving mechanism starts from a brushed DC permanent magnet motor, whose shaft is coupled with multiple gear trains. In between, there is a dry plate clutch that slips to protect the motor when the required actuation torque exceeds the motor's rating. The last gear stage is then coupled with a bevel pinion and gear, which changes the axis of rotation. Its output side will actuate the drive bearing located in the escapement of a drive bar. This mechanism allows the switch rails to be unlocked first. Then, the switch rail actuation will be initiated when the lock nuts on the drive rod make contact with the drive lug on the drive stretcher bar, as shown in Figure 5-3.

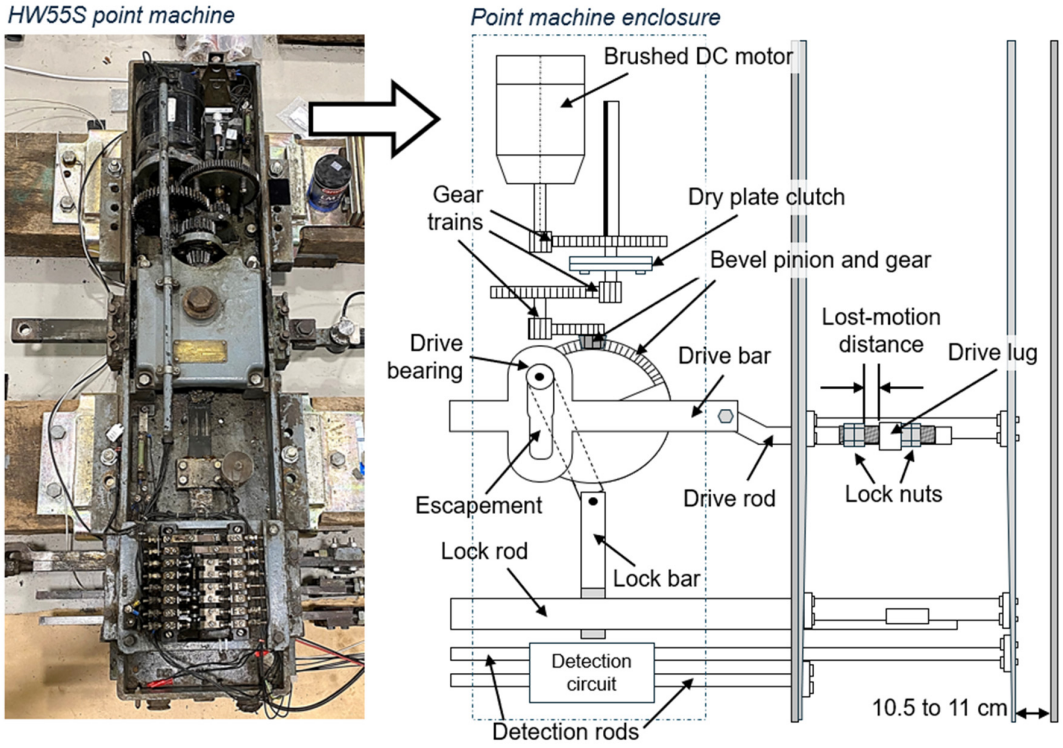


Figure 5-3 Structure and components of the HW55S turnout system

Figure 5-3 displays the state when the HW55S turnout is rested in the "normal" position. The mechanism of the switch rail's actuation from the "normal" to the "reverse" position is illustrated in Figure 5-4. The blue arrows indicate the moving direction of the bevel gear, the greens are for the drive bearing, and the reds are for the drive bar and the drive rod. A detailed description of this operation is provided as follows:

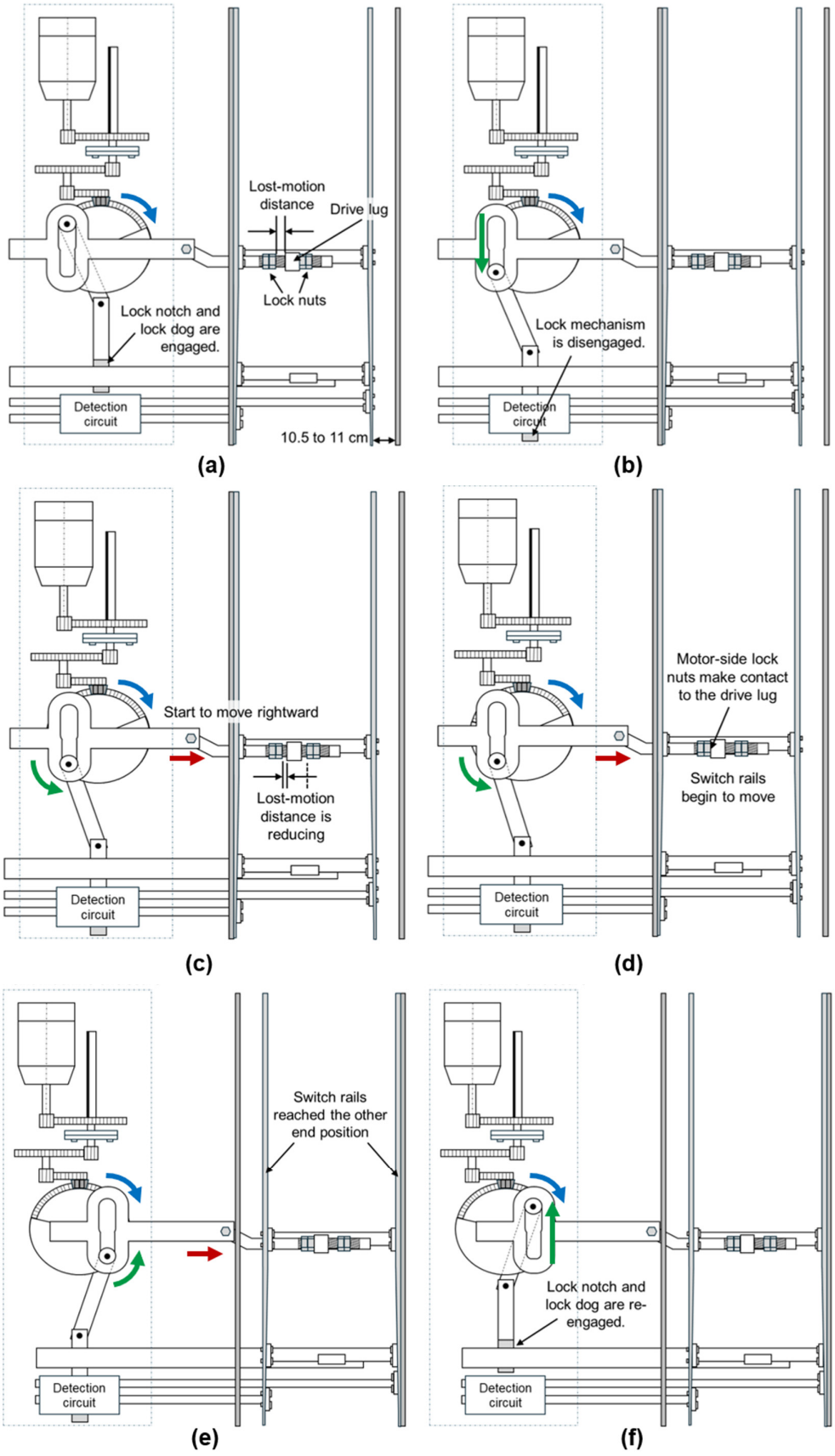


Figure 5-4 Actuation mechanism of the HW55S turnouts operating in N-to-R direction

- When a command of normal-to-reverse (i.e. N-to-R) is sent to the turnout, the drive bearing will first be pushed downward in the escapement. It will also push the lock bar down to disengage the lock notch and the lock dog.
- Once the switch rails are unlocked, the following trajectory of the drive bearing will push the drive bar to the right, which always has a stroke of around 15 cm.
- As the switch rails only needed to be actuated around 10.5 to 11 cm (depending on the initial set-up), the first part of the drive rod stroke will be idle where the switch rails are not yet actuated. This section is called the “lost-motion” distance.
- When the lock nuts make contact with the drive lug, the rail actuation will begin.
- Once the switch rails reach the desired position, the drive bearing will be pulled upward. It will also pull the lock bar back up to re-engage the lock notch and the lock dog, securing the rails. For R-to-N (reverse-to-normal) throws, everything will work identically, except that the directions will all be reversed.

Since the proposed SRIVC-based FDD methodology has only been validated with simulation models in the two previous chapters, its practicality will be validated with measurements obtained from this real HW55S turnout system.

These include the input voltage, motor current, and motor speed of each throw from all simulated health states. An additional linear potentiometer is also installed for observation purposes, especially to inspect the turnout behaviour when faults affecting the rail actuation stroke are simulated. Details and locations of sensor installation are listed in Table 5-1 and illustrated in Figure 5-5.

Table 5-1 All sensors used in the data acquisition system and their details

Sensor types	***Measured variables***	***Part numbers and manufacturers***	***Install locations***
Voltage transducer	Input voltage	LEM LV 100	Across the power supply's terminals
Current transducer	Motor armature current	LEM HTR 50-SB	Enclosing one of the cables from the DC power supply
Rotary encoder	Motor shaft speed	Kubler 05.2400.1122.0100	At the end of the first gear stage shaft
Linear potentiometer	Switch rails moving displacement	Curtiss-wright SLS190/250/10K/L/66/1	Fix one end on a base and the other to the lock stretcher bar

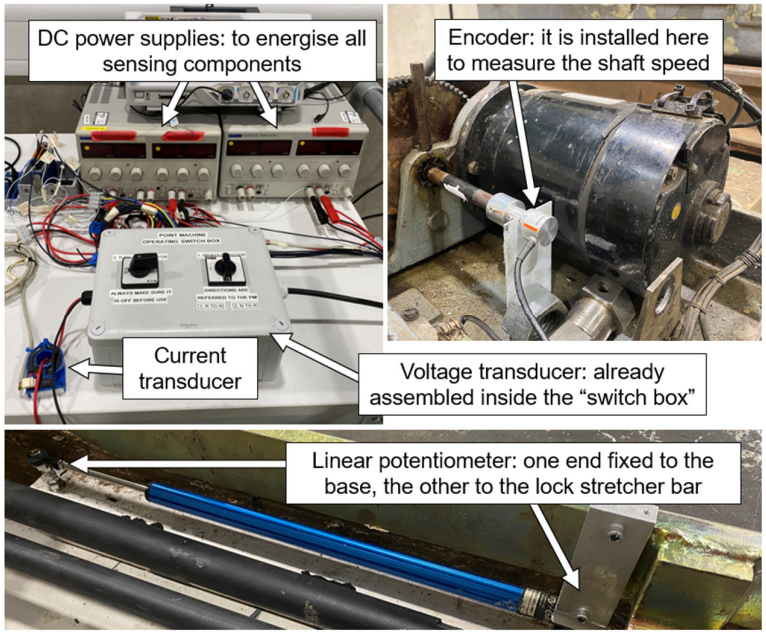


Figure 5-5 All sensor installations in the HW55S turnout system for data acquisition

Readings from every sensor are measured and collected by a data acquisition system consisting of an NI cDAQ-9174 chassis, an NI-9205 voltage sensing module, and software written in the MATLAB® application running on a personal laptop.

Since the maximum aggregated sampling rate of this module is 250 kSample/s, 25 kHz is selected for all four channels for this data acquisition. This sampling rate has already been validated with turnout simulation scenarios, which performed parameter change detection effectively and did not require a long computational time.

When the data acquisition campaign is commenced, the HW55S turnout will be operated alternatively between two directions, normal-to-reverse (N-to-R) and reverse-to-normal (R-to-N), by using the “switch box” located on the table (see Figure 5-5). The practicality of the proposed methodology will be evaluated in the next section when it first focuses on fault-free conditions since their estimates are required to generate the baseline of each parameter for the subsequent parameter change detection.

5.2 Fault-free conditions of HW55S turnout

Once the HW55S turnout has been assembled, measurements of when it runs in fault-free conditions will be collected. Then, their SRIVC parameter estimates will be obtained to generate the baseline that represents the fault-free characteristics of each turnout parameter. Figure 5-6 displays motor current and motor speed measurements from 206 fault-free N-to-R and 210 fault-free R-to-N throws when a constant DC power supply of 110 V energises the HW55S turnout up until the end of the throw.

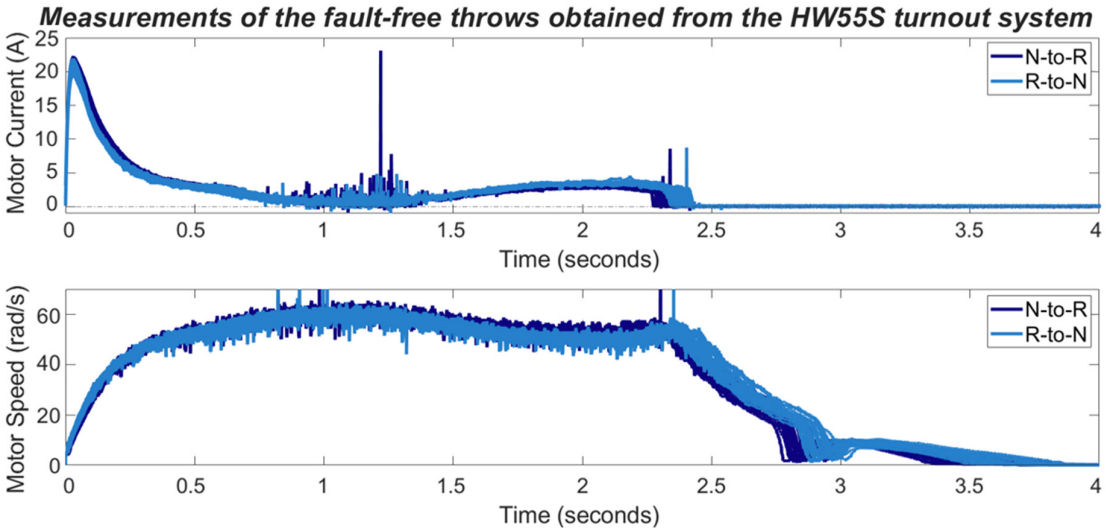


Figure 5-6 Measurements obtained from fault-free throws of the HW55S turnout system

Important features of these measurements are highlighted as follows:

- Asymmetrical mechanisms of the HW55S turnout caused the measurements of fault-free N-to-R and R-to-N throws to be slightly different.

- Measurements of the same direction of throw also show variations, which are the results of both noises and random differences between manoeuvres. As a result, ***the fault-free condition is not deterministic, but there is a range of conditions that should still be considered fault-free***.

Compared to the estimation of the turnout simulation models, the "true" model structures of the turnout and statistics of measurement noises are now completely unknown. Nevertheless, the same transfer function models, i.e. Eqs. (3-29) and (3-42), which have been validated in Chapter 4, will still be used for the following estimation.

To verify the practicality of the SRIVC algorithm with the measurements from a real HW55S turnout system, it was applied to one of the datasets plotted in Figure 5-6 with the validated transfer function models. The SRIVC parameter estimation results of the selected fault-free throw are shown in Figure 5-7 and Table 5-2.

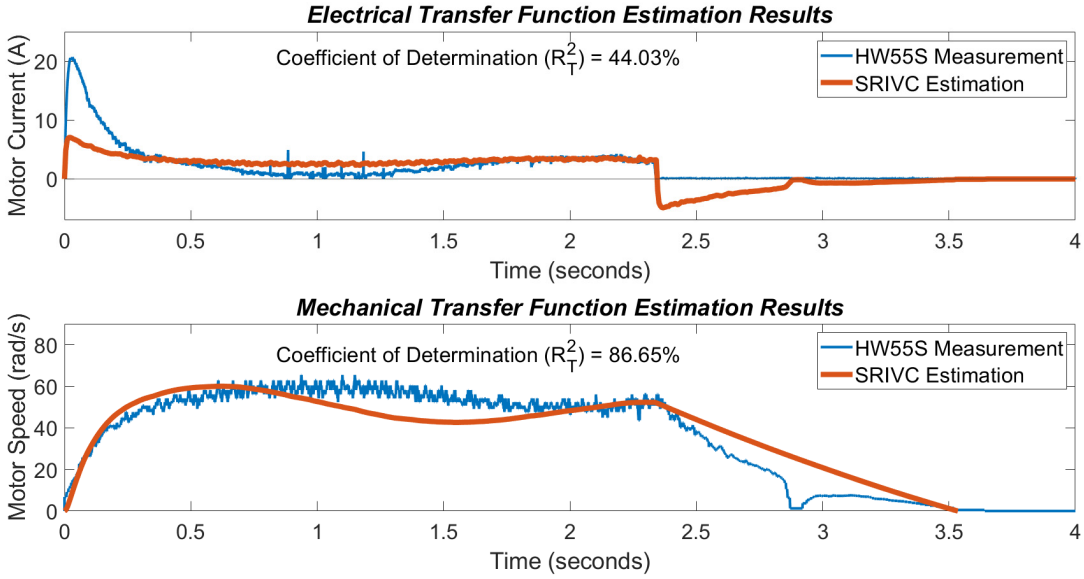


Figure 5-7 Comparison between the fault-free measurements from the selected throw and the responses of the validated transfer functions with the SRIVC parameter estimates

Table 5-2 SRIVC estimation results of the selected throw and the numbers in parentheses as their corresponding estimated standard deviation

Electrical Parameter	***Estimation result***	***Mechanical Parameter***	***Estimation result***
R_A – Resistance (Ω)	13.603 (1.119)	J_{TO} – Turnout's Inertia (Nm/(rad/s^2))	0.0690 (0.0057)
L_A – Inductance (mH)	66.60 (2.722)	B_{TO} – Turnout's Damping (Nm/(rad/s))	0.0331 (0.0029)
k_M – Constants (V/(rad/s) and Nm/A)	1.2736 (0.1038)	K_{TO} – Turnout's Stiffness (Nm/rad)	0.0175 (0.0015)

Figure 5-7 compares the measurements obtained from the HW55S turnout and the responses of the validated transfer functions with the SRIVC parameter estimates. The coefficient of determination (R_T^2) is also used to assess data-fitting performance, which indicates that although the turnout parameter estimates are positive and have low estimated standard deviations, R_T^2 from the estimation of the electrical transfer function is less than 80% (the same threshold used in Chapter 3 for the validity test).

This poor performance is the result of applying the SRIVC algorithm to the entire throw duration, also known as "en-bloc" parameter estimation, which can only perform satisfactorily when the system dynamics are unchanged. Based on the nature of the turnout driving mechanism, the en-bloc SRIVC algorithm is not suitable for estimating the turnout parameters. The justifications for this argument are elaborated as follows:

- During the inrush and unlocking phases, the load seen by the DC motor is only contributed by mechanical parts inside the point machine. This is when the drive bearing is pushed in the escapement slot to release the engagement between the lock notch and the lock dog to unlock the switch rails. Thus, the interaction between the point machine and the switch rails has yet to happen.
- Once the lock nuts on the drive rod make contact with the drive lug, the switch rail's actuation will then be initiated. This is when the motor sees an increase in

load as it also has to actuate the switch rails during the moving phase. The interaction between the point machine and the switch rails will last until the switch rails are successfully moved, locked, and detected.

Therefore, the motor inside the point machine experiences two different loading phases. The first is when the motor has to overcome the internal mechanical resistance to unlock the switch rails. The second is when it also has to actuate the switch rails and lock them in place. These changes cause the turnout dynamics to be non-linear and cause the en-bloc SRIVC parameter estimation to be underperformed. The next section proposes modifications that will enhance the SRIVC algorithm's practicality.

5.2.1 Modifications of SRIVC parameter estimation

Previous studies have shown that the analysis of non-linear turnout dynamics can be conducted more effectively with the integration of "segmentation" (Roberts et al., 2002; Gao et al., 2022; Guzman et al., 2022; Wang et al., 2023). As stated above, since the motor sees two different loading conditions, the segmentation implicated in this thesis will automatically divide any testing throw into two phases: "inrush + unlock" and "move + lock". Similar to the algorithm proposed by Ou, Xue, and Cui (2019), the proposed segmentation algorithm is developed and outlined as follows:

First, the algorithm requires the users to provide an initial range that covers the instant that the moving phase begins. This transitional point is essential as it is when the DC motor starts to experience different loads. Once the switch rail's actuation is initiated, the required torque from the motor will be increased, causing its speed to drop and the current to rise. Therefore, these trend changes could facilitate the initial range determination. Based on the measurements in Figure 5-6, the transitional point could be somewhere between 0.8 and 1.4 seconds.

From the beginning of the throw to just before 0.8 seconds, the motor speed always increases. It is reduced considerably when the switch rail's actuation is initiated. Since the trend of the motor speed changes continuously from rising to falling, there will be an instant where the speed is constant and produces the first derivative of zero.

A moving-average filter with a window of 2,500 samples (i.e. 10% of the data acquisition sampling rate) is implemented to reduce the high-frequency noises in the speed measurements. Then, their first derivatives are approximated using the "`diff`" routine in MATLAB®. Following this, another identical filtering step is applied to the first derivatives of the motor speed to reduce the differentiation errors. Since the instant where the first derivative is exactly zero may not occur in realistic measurements, the minimum absolute value of the filtered first derivatives will be chosen as the transitional point instead. All of these segmentation procedures are illustrated in Figure 5-8.

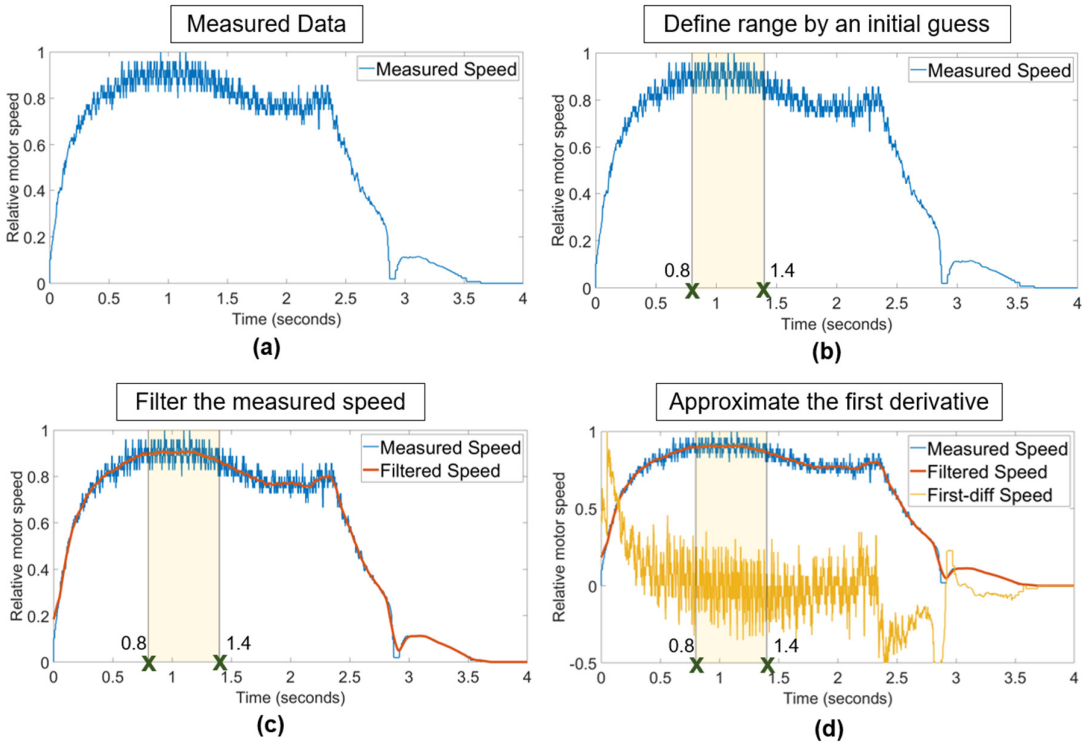


Figure 5-8 Procedural steps of the proposed segmentation algorithm

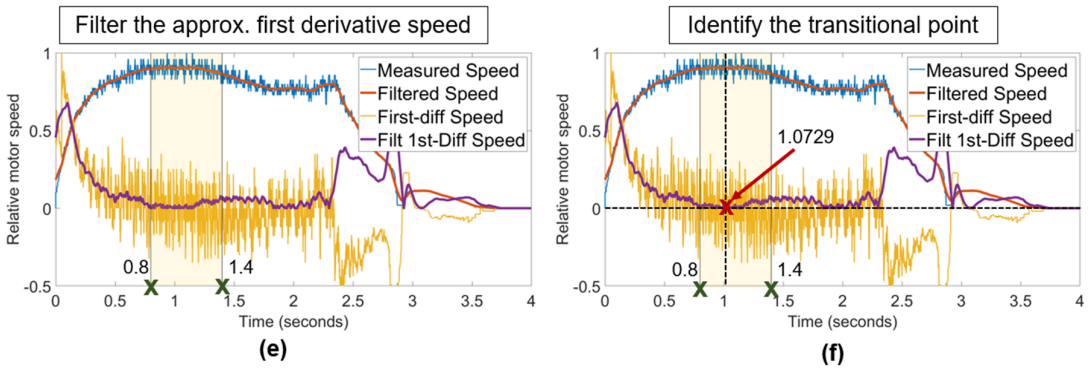


Figure 5-8 (cont.) Procedural steps of the proposed segmentation algorithm

The developed segmentation algorithm is also found to be sufficiently robust to the initial range selection. For any different initial range that contains the transitional phase of the motor speed from rising to falling, the segmentation always produces the same transitional time, showing the relaxation of the expert system knowledge.

The purpose of the segmentation algorithm is to separate the non-linear turnout dynamics into two linear phases. Since the “inrush + unlock” phase begins when the turnout is at rest, its initial conditions are all zero. Meanwhile, the latter “move + lock” phase starts after 1.0729 seconds, so its initial conditions will not be zero. These non-zero initial conditions require the transfer function models to be modified first before the estimation (Garnier et al., 2003; Padilla, 2017), which can be derived as follows.

First, the electrical differential equation is repeated here for convenience:

$$v_s(t) = L_A \frac{di_a(t)}{dt} + R_A i_a(t) + k_M \omega_m(t) \tag{3-27}$$

The Laplace transform of this equation when $i_a(t_0) \neq 0$ and $\omega_m(t_0) \neq 0$ is:

$$V_s(s) = L_A\left[sI_a(s) - i_a(t_0)\right] + R_A I_a(s) + k_M \Omega_m(s) \tag{5-1a}$$

$$\left[\frac{1}{sL_A + R_A}\right] V_s(s) = \left[\frac{1}{sL_A + R_A}\right](sL_A + R_A) I_a(s) + \left[\frac{1}{sL_A + R_A}\right] k_M \Omega_m(s) + \left[\frac{1}{sL_A + R_A}\right]\left[-L_A i_a(t_0)\right] \tag{5-1b}$$

$$sI_{a,f}(s) = \left(\frac{-R_A}{L_A}\right) I_{a,f}(s) + \left(\frac{1}{L_A}\right) V_{s,f}(s) + \left(\frac{-k_M}{L_A}\right) \Omega_{m,f}(s) + \left[\frac{i_a(t_0)}{sL_A + R_A}\right] \quad \text{(5-1c)}$$

Transforming this equation into the time-domain linear regression format yields:

$$\frac{di_{a,f}(t)}{dt} = \left(\frac{-R_A}{L_A}\right) i_{a,f}(t) + \left(\frac{1}{L_A}\right) v_{s,f}(t) + \left(\frac{-k_M}{L_A}\right) \omega_{m,f}(t) + \left[i_a(t_0)\right] h_{e,f}(t) \quad \text{(5-2)}$$

where $h_{e,f}(t)$ is the impulse response of the electrical pre-filter (i.e. $1/(sL_A + R_A)$).

Second, the mechanical differential equation is repeated here for convenience:

$$k_T i_a(t) = J_{TO} \frac{d\omega_m(t)}{dt} + B_{TO}\omega_m(t) + K_{TO}\int_0^t \omega_m(\tau)\, d\tau \quad \text{(3-40)}$$

The above equation only holds if the switch rails interact with the point machine immediately after the turnout is activated, which was assumed for the previous turnout simulation scenarios. However, for a real HW55S turnout, these two components have yet to engage during the “inrush + unlock” phase. Therefore, the mechanical differential equation in Eq. (3-40) must be modified as follows:

$$k_T i_a(t) = J_{PM} \frac{d\omega_m(t)}{dt} + B_{PM}\omega_m(t); \text{ since } K_{TO} \text{ is assumed to be zero} \quad \text{(5-3)}$$

$$\frac{d\omega_m(t)}{dt} = \left(\frac{-B_{PM}}{J_{PM}}\right) \omega_m(t) + \left(\frac{k_T}{J_{PM}}\right) i_a(t) \quad \text{(5-4)}$$

This modification is implemented since it is still assumed that all of the stiffness effect only comes from the switch rails. Furthermore, the subscripts of the mechanical parameters are now changed to emphasise that they are contributed only by the point machine, i.e. its inertia (J_{PM}) and its damping (B_{PM}).

Then, during the “move + lock” phase, the switch rails have engaged with the point machine. The switch rails’ characteristics, including their stiffness effect, will now be seen by the motor. Thus, once the entire Eq. (3-40) is differentiated to eliminate the integration term, its Laplace transform when $i_a(t_0) \neq 0$ and $\omega_m(t_0) \neq 0$ is:

$$k_T \frac{di_a(t)}{dt} = J_{TO} \frac{d^2\omega_m(t)}{dt^2} + B_{TO} \frac{d\omega_m(t)}{dt} + K_{TO}\omega_m(t) \quad (5\text{-}5a)$$

$$sk_T I_a(s) = \left[s^2 J_{TO} + sB_{TO} + K_{TO} \right] \Omega_m(s) + k_T i_a(t_0) - sJ_{TO}\omega_m(t_0) - J_{TO} \frac{d\omega_m}{dt}(t_0) - B_{TO}\omega_m(t_0) \quad (5\text{-}5b)$$

$$\left[\frac{1}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] sk_T I_a(s) = \left[\frac{1}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] \left[s^2 J_{TO} + sB_{TO} + K_{TO} \right] \Omega_m(s) + \left[\frac{-sJ_{TO}\omega_m(t_0)}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] + \left[\frac{k_T i_a(t_0) - J_{TO}\dot{\omega}_m(t_0) - B_{TO}\omega_m(t_0)}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] \quad (5\text{-}5c)$$

$$sk_T I_{a,f}(s) = \left[s^2 J_{TO} + sB_{TO} + K_{TO} \right] \Omega_{m,f}(s) + \left[\frac{-sJ_{TO}\omega_m(t_0)}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] + \left[\frac{k_T i_a(t_0) - J_{TO}\dot{\omega}_m(t_0) - B_{TO}\omega_m(t_0)}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] \quad (5\text{-}5d)$$

$$s^2 \Omega_{m,f}(s) = \left(\frac{-B_{TO}}{J_{TO}} \right) s\Omega_{m,f}(s) + \left(\frac{-K_{TO}}{J_{TO}} \right) \Omega_{m,f}(s) + \left(\frac{k_T}{J_{TO}} \right) sI_{a,f}(s) + \left[\frac{s\omega_m(t_0)}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] + \left[\frac{-k_T i_a(t_0) + J_{TO}\dot{\omega}_m(t_0) + B_{TO}\omega_m(t_0)}{J_{TO}\left(s^2 J_{TO} + sB_{TO} + K_{TO} \right)} \right] \quad (5\text{-}5e)$$

Transforming this equation into the time-domain linear regression format yields:

$$\frac{d^2\omega_{m,f}(t)}{dt^2} = \left(\frac{-B_{TO}}{J_{TO}} \right) \frac{d\omega_{m,f}(t)}{dt} + \left(\frac{-K_{TO}}{J_{TO}} \right) \omega_{m,f}(t) + \left(\frac{k_T}{J_{TO}} \right) \frac{di_{a,f}(t)}{dt} + \mathcal{L}^{-1}\left[\frac{s\omega_m(t_0) + \Psi}{s^2 J_{TO} + sB_{TO} + K_{TO}} \right] \quad (5\text{-}6)$$

where Ψ is the combined initial condition terms. Its only importance is to enforce the correct structure of the last term's numerator, not to imply the occurrence of any fault.

In conclusion, the non-linear turnout dynamics have now been divided into two linear phases with the following associated transfer function models:

Transfer functions describing the dynamics during the "inrush + unlock" phase:

$$I_a(s) = \left(\frac{1}{sL_A + R_A}\right) V_s(s) + \left(\frac{-k_M}{sL_A + R_A}\right) \Omega_m(s) \tag{5-7a}$$

$$\Omega_m(s) = \left[\frac{k_T}{sJ_{PM} + B_{PM}}\right] I_a(s) \tag{5-7b}$$

Transfer functions describing the dynamics during the "move + lock" phase:

$$I_a(s) = \left(\frac{1}{sL_A + R_A}\right) V_s(s) + \left(\frac{-k_M}{sL_A + R_A}\right) \Omega_m(s) + \left(\frac{i_a(t_0)}{sL_A + R_A}\right) \tag{5-8a}$$

$$\Omega_m(s) = \left[\frac{k_T}{s^2 J_{TO} + sB_{TO} + K_{TO}}\right] sI_a(s) + \left[\frac{s\omega_m(t_0) + \Psi}{s^2 J_{TO} + sB_{TO} + K_{TO}}\right] \tag{5-8b}$$

As a consequence, there are ten parameters to be estimated from this proposed methodology. These are three electrical parameters (R_A, L_A, and k_M), two mechanical parameters characterising the point machine (J_{PM} and B_{PM}), another three that include the effects of the switch rails as well (J_{TO}, B_{TO}, and K_{TO}), and two initial conditions ($i_a(t_0)$ and $\omega_m(t_0)$). In the following section, the SRIVC algorithm is re-applied to these modified transfer functions and its parameter estimation performance is re-assessed.

5.2.2 Baseline of all fault-free HW55S turnout parameters

Once the non-linear turnout dynamics have been divided into two linear phases and their associated transfer function models are derived, i.e. Eqs. (5-7) and (5-8), the estimation results of the same selected fault-free throw (see Figure 5-7) can be listed in Table 5-3. Meanwhile, the data-fitting performance of this proposed methodology is also displayed in Figure 5-9, with the results of both phases being separated for clarity.

Table 5-3 SRIVC parameter estimates of the same selected throw, as plotted in Figure 5-7, with their corresponding estimated standard deviation in parentheses

Turnout Parameters	***Estimation result***	***Turnout Parameters***	***Estimation result***
R_A – Resistance (Ω)	4.5423 (0.0451)	J_{TO} – Turnout's Inertia (Nm/(rad/s^2))	0.0713 (0.0009)
L_A – Inductance (mH)	38.06 (0.1867)	B_{TO} – Turnout's Damping (Nm/(rad/s))	0.1786 (0.0026)
k_M – Constants (V/(rad/s) and Nm/A)	1.8266 (0.0179)	K_{TO} – Turnout's Stiffness (Nm/rad)	0.0943 (0.0014)
J_{PM} – Point machine's Inertia (Nm/(rad/s^2))	0.1157 (0.0012)	$i_a(t_0)$ – Initial current entering the moving phase (A)	-0.606 (0.0150)
B_{PM} – Point machine's Damping (Nm/(rad/s))	0.0300 (0.0005)	$\omega_m(t_0)$ – Initial speed entering the moving phase (rad/s)	63.362 (0.0623)

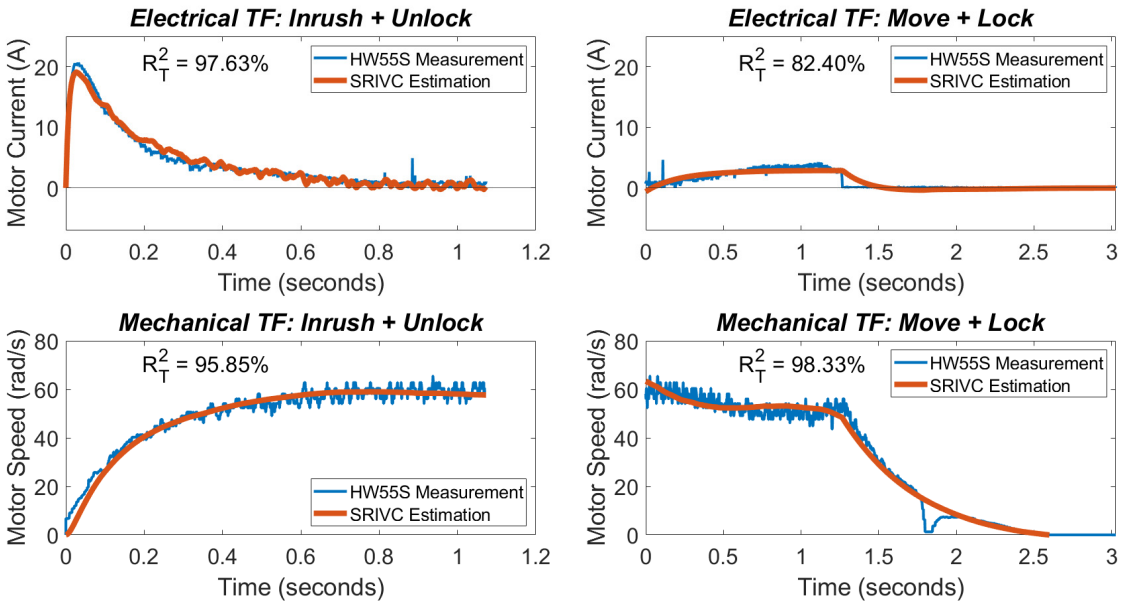


Figure 5-9 Comparison between the measurements of the same selected throw, as shown in Figure 5-7, and the results from the estimation

The results in Figure 5-9 demonstrate that three out of four estimations produce a very high coefficient of determination (more than 95%). The worst is the estimation of the electrical transfer function during the "move + lock" phase. This is the result of an open circuit happening around 1.25 seconds after the moving phase has begun, which corresponds to when the switch rails have reached the end position. This circuit isolation creates another non-linearity that was not accounted for.

Therefore, the estimation results from the other three transfer functions will be the only focus from this point onwards. This is justifiable because abrupt faults affecting the turnout's electrical aspects are not in this thesis's scope, making the initial current entering the moving phase (i.e. $i_a(t_0)$) the only added information from the estimation of the "move + lock" electrical transfer function. Due to the non-linearity from an open circuit, it was also incorrectly estimated and had a negative value. As a consequence, instead of this parameter, the total displacement of the switch rail's actuation (*TotDist*) measured from the additional linear potentiometer will be considered instead. For this specific fault-free throw, the switch rails were moved 10.96 cm.

The methodology has been modified, so the baseline for each parameter can now be formulated by applying parameter estimation to all fault-free throws, shown in Figure 5-6. Since not every experiment was conducted on the same day, the HW55S turnout's characteristics may have changed unknowingly. To eliminate the effects of such uncontrollable variations, the parameter estimates are converted to the relative differences based on the interquartile range (IQR) of that parameter calculated from the measurements obtained on the same day. Therefore, the general equation for the normalisation for each of the turnout parameter is given by Eq. (5-9):

$$\hat{\rho}^{r}_{i,j} = \frac{\hat{\rho}_{i,j} - IQR_i}{IQR_i} \times 100\% \tag{5-9}$$

where IQR_i is the interquartile range of the i^{th} parameter calculated from the estimates of that parameter obtained within the same day;

$\hat{\rho}_{i,j}$ is the raw numerical estimate of the i^{th} parameter of the j^{th} operation, where $\hat{\rho}$ is the estimation of [R_A, L_A, k_M, J_{PM}, B_{PM}, J_{TO}, B_{TO}, K_{TO}, $\omega_m(t_0)$, *TotDist*];

$\hat{\rho}^{r}_{i,j}$ is the relative difference value of the i^{th} parameter of the j^{th} operation.

Thus, the estimation results of all fault-free operations (see Figure 5-6) are presented by a series of boxplots in Figure 5-10. It contains two sets of ten boxplots, each representing R_A, L_A, k_M, J_{PM}, B_{PM}, J_{TO}, B_{TO}, K_{TO}, $\omega_m(t_0)$, and *TotDist*, respectively.

The boxplots show that the parameter estimates of all fault-free HW55S throws have their median around 0%, with fairly symmetrical distributions. Most estimates also have a standard deviation of less than 20%, with the exception of B_{PM} and K_{TO} that are widely distributed. These would later cause them to be less significant in the later FDD analysis. Finally, even if the distributions of N-to-R and R-to-N throws are similar, they are not identical, which reaffirms the asymmetry of the HW55S turnout system.

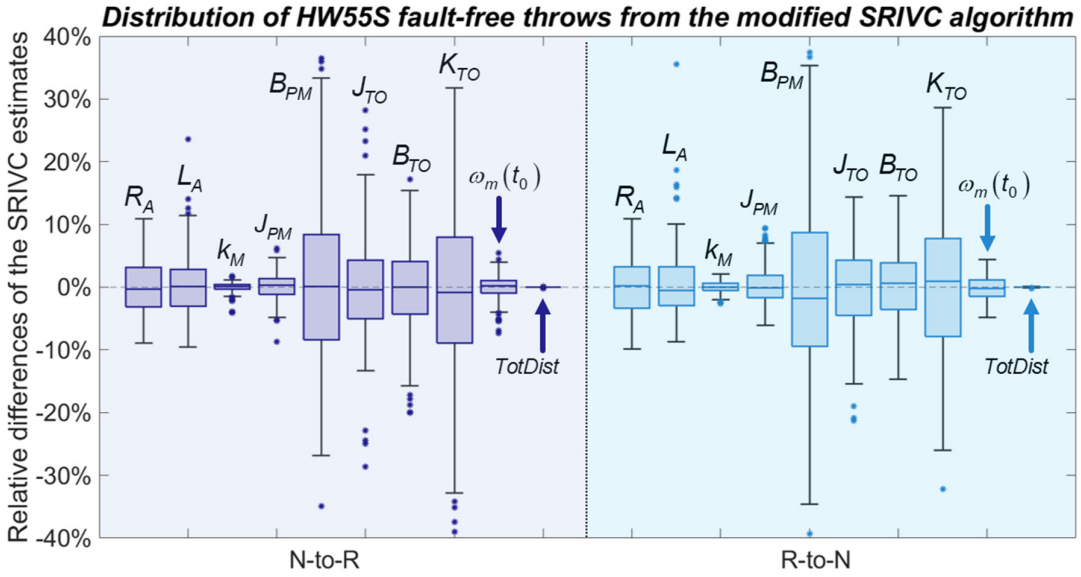


Figure 5-10 Distributions of parameter estimates from all fault-free HW55S throws using the interquartile range as the baseline for each parameter

The boxplots in Figure 5-10 prove that the interquartile range of each parameter from fault-free throws is suitable to be used as the baseline. This concept is applied in the following section to calculate the relative differences when the measurements are now obtained from the simulations of five fault modes. The relative differences that are outstanding for each mode are the main FDD features, which facilitate the creation of the knowledge-base table at the end of this chapter.

5.3 Fault simulations imposed onto the HW55S turnout

The fault simulations conducted in this thesis only aim to replicate the effects of each fault by artificially mal-adjusting the HW55S turnout. This process is convenient for data acquisition as data labelling can be done in a controllable and deterministic manner. Details of each fault simulation and their estimation results are elaborated in the following sub-sections. The figures listed in Table 5-4 also provide the number of throws simulated for each fault mode, including the fault-free one. These are regarded as the "training" dataset for the subsequent FDD algorithm development.

Table 5-4 Number of throws simulated during the training data collection campaign

Fault modes	***Directions***	***Number of operations***		
		Fault level 1	*Fault level 2*	*Fault level 3*
Fault-free	N-to-R	206		
	R-to-N	210		
Overdrive	N-to-R	116	39	117
	R-to-N	117	88	120
Unlock difficulty	N-to-R	116	88	118
	R-to-N	117	39	117
Shim obstruction	N-to-R	128	48	124
	R-to-N	128	44	127
OPP-obstruction* (page 148)	N-to-R	128	45	128
	R-to-N	130	49	121
Underdrive	N-to-R	71	91	71
	R-to-N	69	59	73
OPP-underdrive* (page 154)	N-to-R	69	58	74
	R-to-N	71	96	72
Abnormal input voltage	N-to-R	109	78	119
	R-to-N	109	80	120
Improper sliding surface	N-to-R	111	243	(N/A)
	R-to-N	110	244	(N/A)
Total in each direction	N-to-R	2,495		
	R-to-N	2,509		

NOTE: Fault conditions with an asterisk will be defined in their respective sub-sections.

5.3.1 Overdrive fault simulations

Overdrive faults are normally caused by misalignments and out-of-adjustments, which happen from human errors during installation, commissioning, and maintenance. Surrounding vibrations from nearby traffic during services can also incrementally move the lock nuts out of their nominal position. As a result, in this thesis, the simulations of this fault will be done by manually relocating the lock nuts, see Figure 5-11.

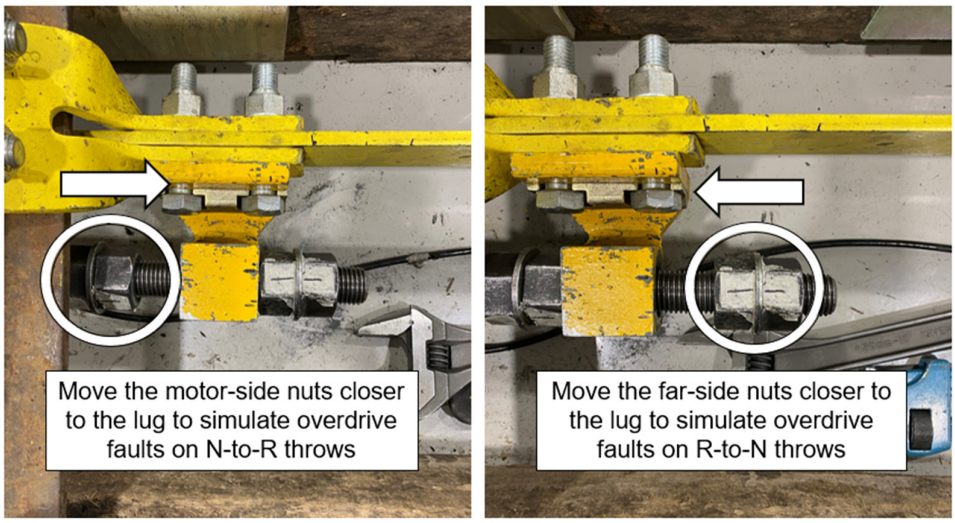


Figure 5-11 Simulation process of overdrive faults conducted on the HW55S turnout

As already described in detail in Section 2.1.4.3, when the lock nuts are located closer to the drive lug than their nominal position, the “lost-motion” distance will be less than the designed value. Since the entire stroke of the drive rod is fixed, the reduction in the “lost-motion” distance will extend the rail actuation stroke. This extension results in a build-up of excessive stress in the switch rails when they are hard-pressed against the stock rail further than in the fault-free scenarios. If the overdriving condition is critical, the increase in internal stress can potentially damage the driving and locking mechanisms, resulting in fractures and cracks in the mechanical linkages as well.

In addition to the effects during the end of the throws, due to the design of the HW55S point machine, when overdriving conditions occur in one direction (assumed

to be R-to-N), it will be more difficult to unlock the switch rails for the following N-to-R throw. This relation is described and illustrated in Figure 5-12.

First, Figure 5-12(a) displays the resting state of the switch rails in the “normal” position. The elastic energy from the rail’s stiffness will try to force them back to their equilibrium. However, because of the engagement between the lock notch and the lock dog, the rails will be secured. When the overdriving condition gets more severe, this restoration elastic force will be much higher, causing the unlocking process to be more difficult; see Figure 5-12(b). Therefore, during the unlocking phase of this N-to-R throw, the DC motor has to produce more force to push the lock dog out of the engagement to unlock the switch rails. This fault condition is commonly known as “unlock difficulty”.

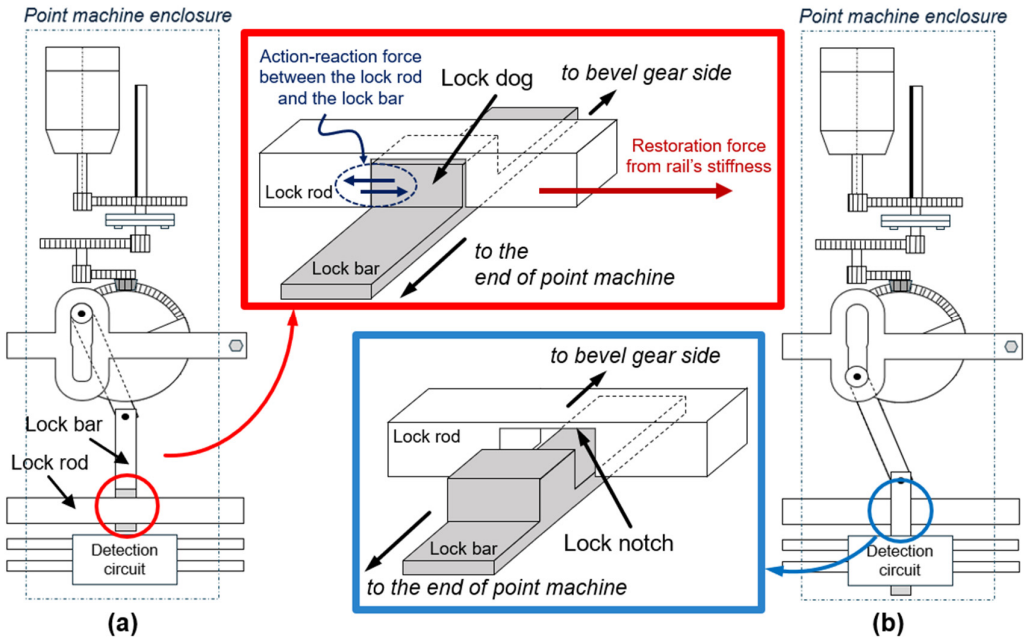


Figure 5-12 Details of HW55S unlocking mechanism by de-coupling the lock dog from the lock notch and the restoration force that makes unlocking more difficult

In this experiment, four severity levels are simulated (including three fault levels and the fault-free condition). The levels are separated based on how many faces the hexagonal lock nuts have been turned to move closer to the drive lug. The three levels are “3-face” (half turn), “6-face” (one turn), and “9-face” (one and a half turns).

For brevity, only the results from N-to-R throws of every fault simulation will be elaborated, since the results from R-to-N throws are comparable. Therefore, they are not presented in the main thesis but are in Appendix 2. The two following sub-sections present the estimation results when overdrive and unlock difficulty faults are simulated on the HW55S turnout system, respectively.

5.3.1.1 *Overdrive faults: estimation results*

The simulation of N-to-R overdrive faults requires maladjustment of the motor-side lock nuts to be closer to the drive lug. After the fault simulations of all four levels, the comparison of measurements from every condition is shown in Figure 5-13.

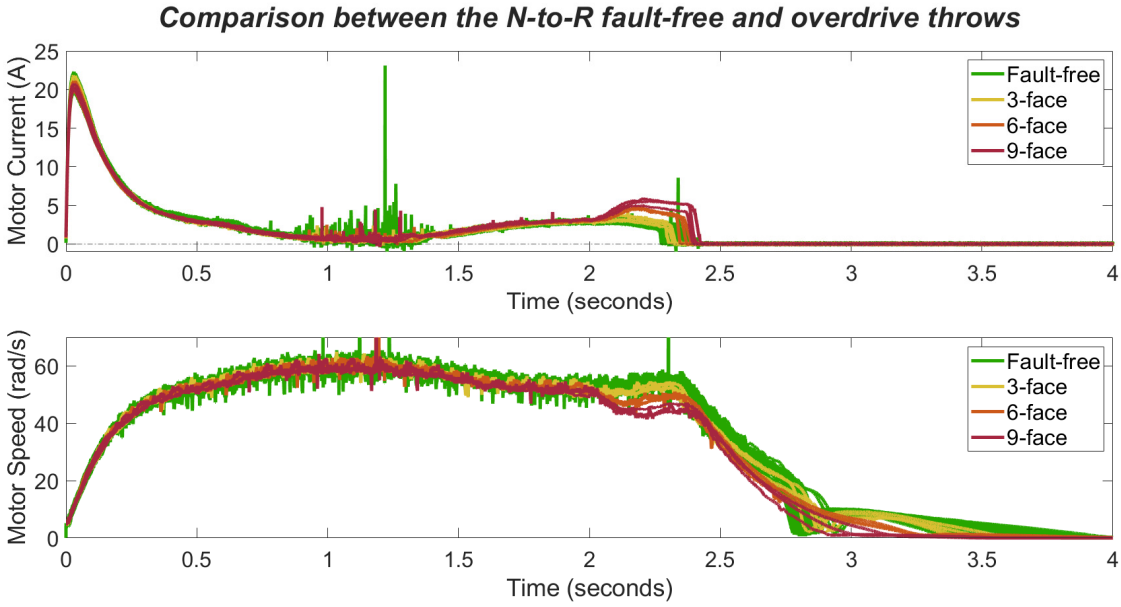


Figure 5-13 Comparison of motor current and motor speed measurements between fault-free and simulated overdrive N-to-R throws

The graphs indicate that the deviations from the fault-free conditions only occur after 2 seconds in both motor current and motor speed, which can be regarded as the beginning of the locking phase. From this instant, the motor current rises gradually and remains constant briefly before suddenly dropping to zero due to an open circuit. The

motor speed, however, declines as the locking becomes more resistive. So, the more severe the overdrive faults simulated, the larger the current rise and the speed drop.

The estimation results of all N-to-R overdrive throws are shown in Figure 5-14 by using the same vertical boxplot format. That is, four groups of ten boxplots are displayed to represent the parameter estimates of each severity level, which is similar to the presentation format applied in Figure 5-10.

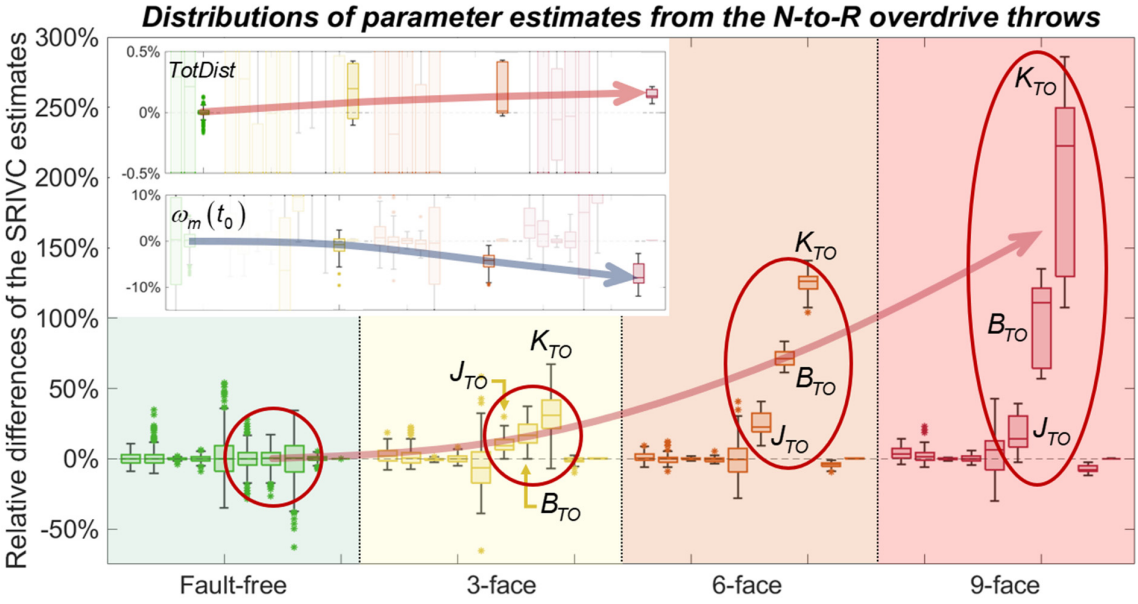


Figure 5-14 Estimation results of overdrive N-to-R throws simulated in all three severity levels, with the zoom-in distributions of TotDist and $\omega_m(t_0)$

The boxplots clearly demonstrate that when more severe overdrive faults are simulated, all mechanical parameters in the "move + lock" phase (i.e. J_{TO}, B_{TO}, and K_{TO}) gain more positive relative difference values. This was expected since overdrive faults increase the mechanical resistance at the end of the throw. Even if the estimates of K_{TO} were previously shown to be less reliable features than the others, the increases in both B_{TO} and K_{TO} between severity levels are clearly distinguished from each other. Such characteristics make them dominant FDD features for the overdrive fault.

Moreover, the zoom-in of *TotDist* and $\omega_m(t_0)$ estimates also shows a clear trend of changes. Since the overdrive throws continue to hard-press the switch rails further against the stock rail, the total switch rail displacement will be slightly more than that in fault-free scenarios. Lastly, because linear transfer function models are employed, the reduction of the speed at the end of the throw will also influence the estimates of the initial motor speed entering the moving phase, i.e. $\omega_m(t_0)$. Therefore, these two parameters could be considered relevant FDD features for overdrive fault diagnosis.

5.3.1.2 *Unlock difficulty faults: estimation results*

In this case, the simulation of N-to-R unlock difficulty faults requires adjustment of the far-side lock nuts to be closer to the drive lug instead (opposite to the previous condition). Similarly, the measurement comparison is illustrated in Figure 5-15.

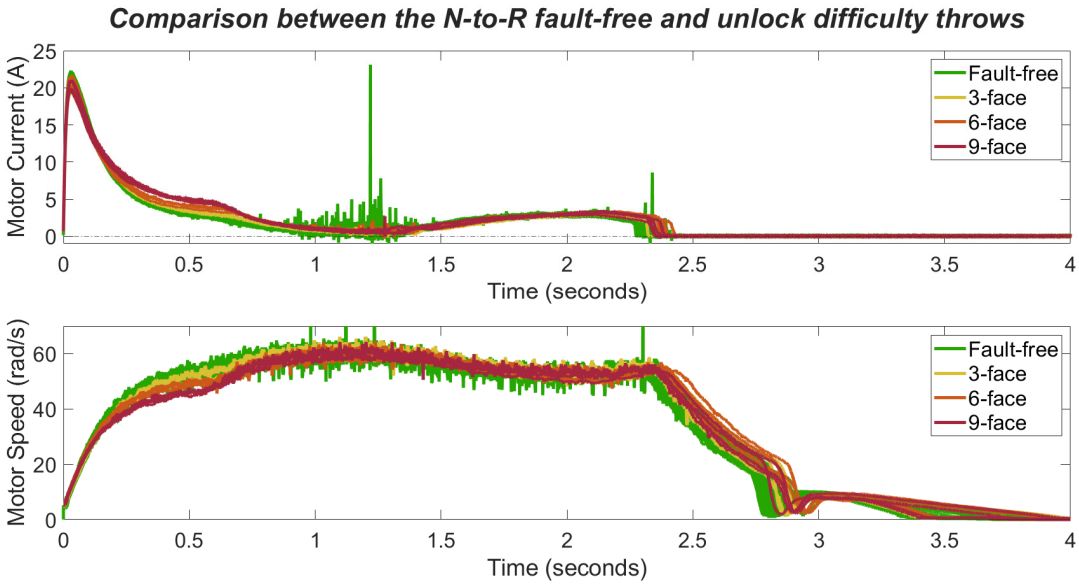


Figure 5-15 Comparison of motor current and motor speed measurements between fault-free and simulated unlock difficulty N-to-R throws

As the name implies, their effects only exist during the unlocking phase of the throws. From around 0.25 seconds onwards, the current decreases at differently rates depending on the severity level. For example, in the “9-face” scenarios, the current

declines more slowly than the less severe and fault-free cases. It settles at a higher level (around 5 A) between 0.5 and 0.7 seconds. This is significantly higher than the fault-free current level during this part of the unlocking phase and corresponds to a corresponding decrease in motor speed. This indicates that the motor must generate a greater actuation torque than usual to release and unlock the switch rails.

Once the switch rails are successfully unlocked (i.e. approximately at 1 second), the waveforms in both motor current and speed will revert to the fault-free conditions. The estimation results of all N-to-R unlock difficulty throws are illustrated in Figure 5-16 with the same vertical boxplot presentation format.

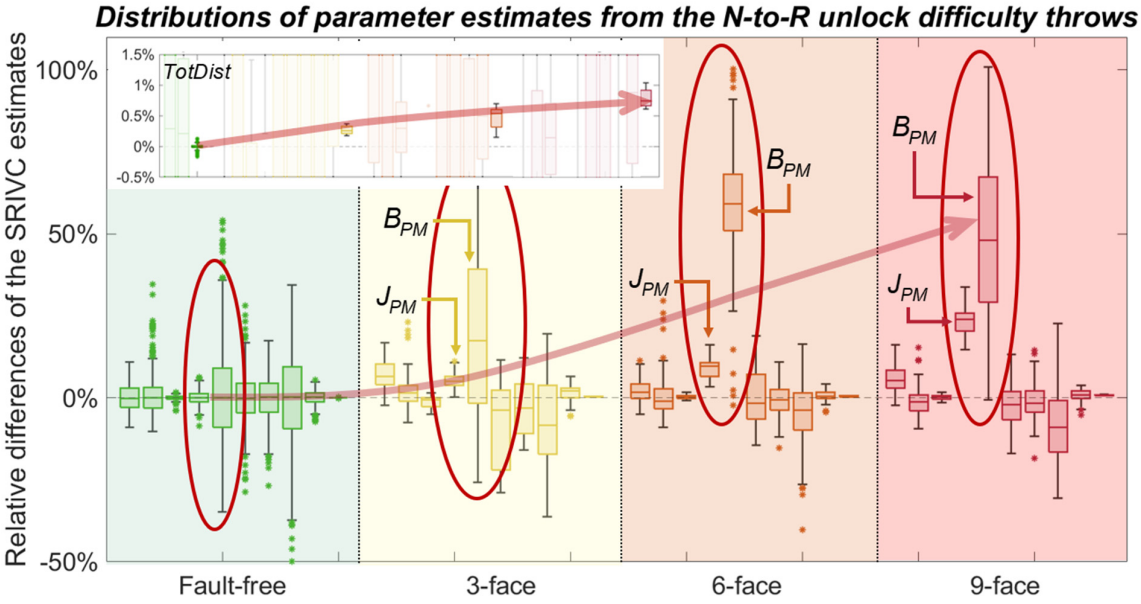


Figure 5-16 Estimation results of unlock difficulty N-to-R throws simulated in all three severity levels, with the zoom-in distributions of TotDist

The boxplots show that only the mechanical parameters in the “inrush + unlock” phase (J_{PM} and B_{PM}) are affected by unlock difficulty faults. The magnitude of changes in B_{PM} is more significant than that of J_{PM}. However, due to the estimates of B_{PM} being widely distributed, they are highly overlapped in all four severity levels. In contrast, even if the magnitude of changes in J_{PM} is lower, their distributions between levels are

better separated. So, changes in both J_{PM} and B_{PM} can imply the occurrence of unlock difficulty fault, but only the magnitude of changes in J_{PM} is indicative of its severity.

Additionally, since the unlock difficulty condition occurs when the other direction of the throw is overdriven, the total switch rail displacement of unlock difficulty throws will have a similar increasing trend. Therefore, the zoom-in of *TotDist* distributions also demonstrates small positive relative values for every unlock difficulty throw.

5.3.2 Obstruction fault simulations

When the HW55S turnout is at rest, there is always an opening gap between one of the switch rails and the stock rail. Thus, there is a chance that obstructions may fall in the gap and block the following throw. Common obstructions range from ballast waste, accumulated and compressed small particles (e.g. coal dust, sand, and ice), litter thrown out from passing trains, and other objects being left by trespassers.

The severity of faults will depend on the size of these obstructions after being compressed. This fault will turn into a failure if the size of the obstacle is larger than the allowance gap of the locking mechanism, as previously shown in Section 2.1.4.1. Meanwhile, if the obstruction is thinner, like a film of compressed dust, leaves, or snow, it will affect the turnout similarly to the overdrive faults. The reason is that the point machine will still be able to actuate the switch rails and bring the lock dog back to the locking position while compressing the obstructions during the locking phase.

In this thesis, the obstruction faults are simulated by attaching a metal shim of three thicknesses that is not larger than the allowable gap of the HW55S point machine to the stock rail at the positions shown in Figure 5-17.

The effects of this fault on N-to-R throws will exist when the shim is attached to the far-side stock rail. To conduct a similar set of experiments compared to overdrive

fault simulations, the effects when the shim is at the motor-side stock rail on N-to-R throws will also be examined. The latter experiments are expected to produce similar results compared to the unlock difficulty conditions. As the obstruction on the motor-side propagates its effects to the direction that does not experience obstruction faults directly, these throws will be defined as “OPP-obstruction”. The next sections provide the estimation results of N-to-R obstruction and OPP-obstruction throws, respectively.

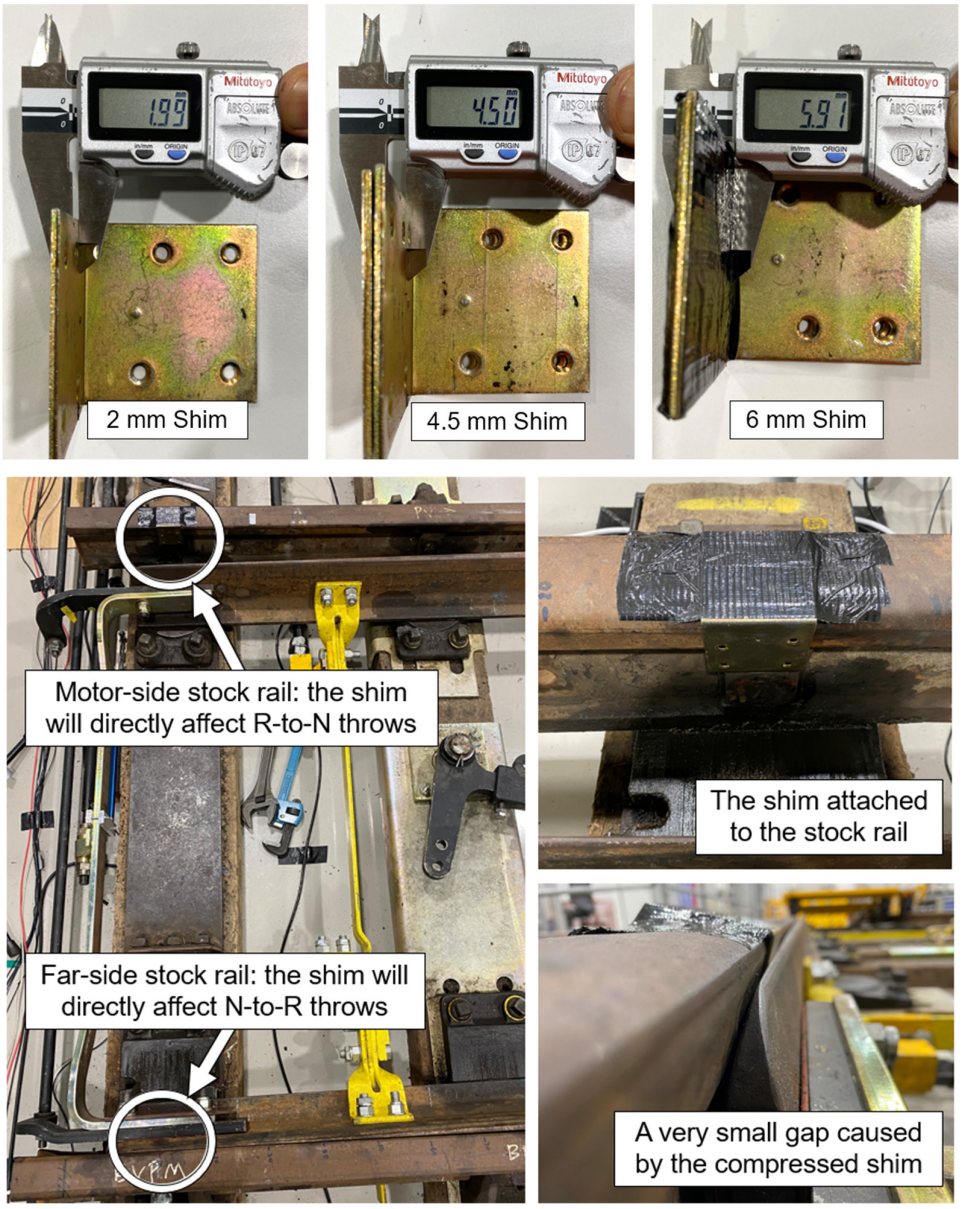


Figure 5-17 Three different thicknesses of metal shim placed in two positions on the HW55S turnout system to simulate obstruction faults

5.3.2.1 Obstruction faults: estimation results

The simulation of N-to-R obstruction faults requires the metal shim to be placed on the far-side stock rail because the switch rails will rest in the “reverse” position after the end of N-to-R throws. Once the simulations of all severity levels are conducted, the measurements from every condition are compared in Figure 5-18.

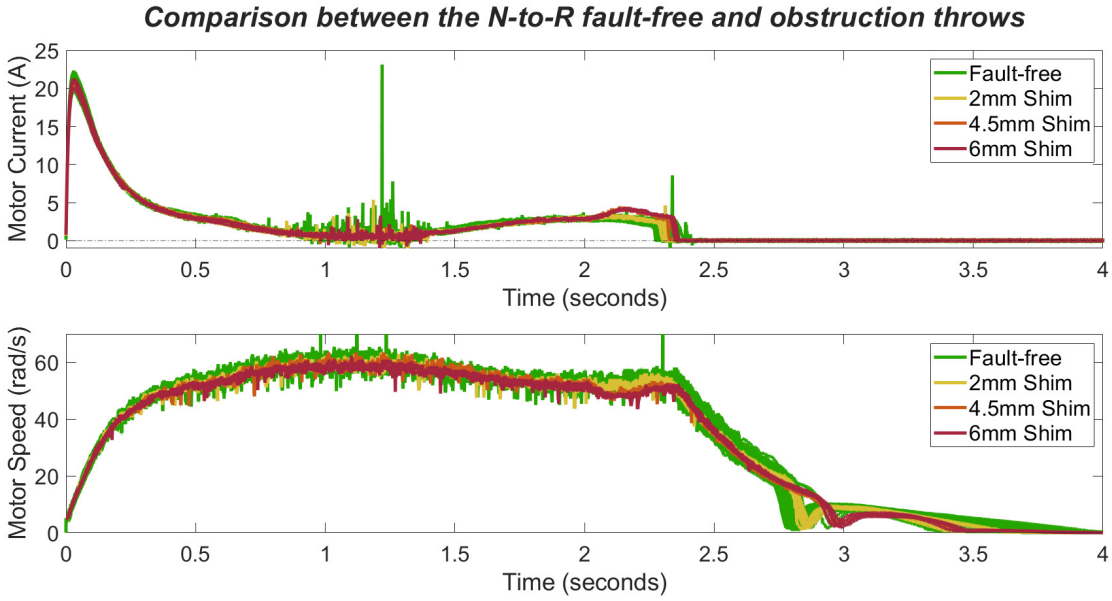


Figure 5-18 Comparison of motor current and motor speed measurements between fault-free and simulated obstruction N-to-R throws

The measurements display many similarities compared to those obtained from the overdrive throws, which are expected. These include the motor current rise and the motor speed drop during the locking phase. The associated estimation results of all N-to-R simulated obstruction throws are provided in Figure 5-19.

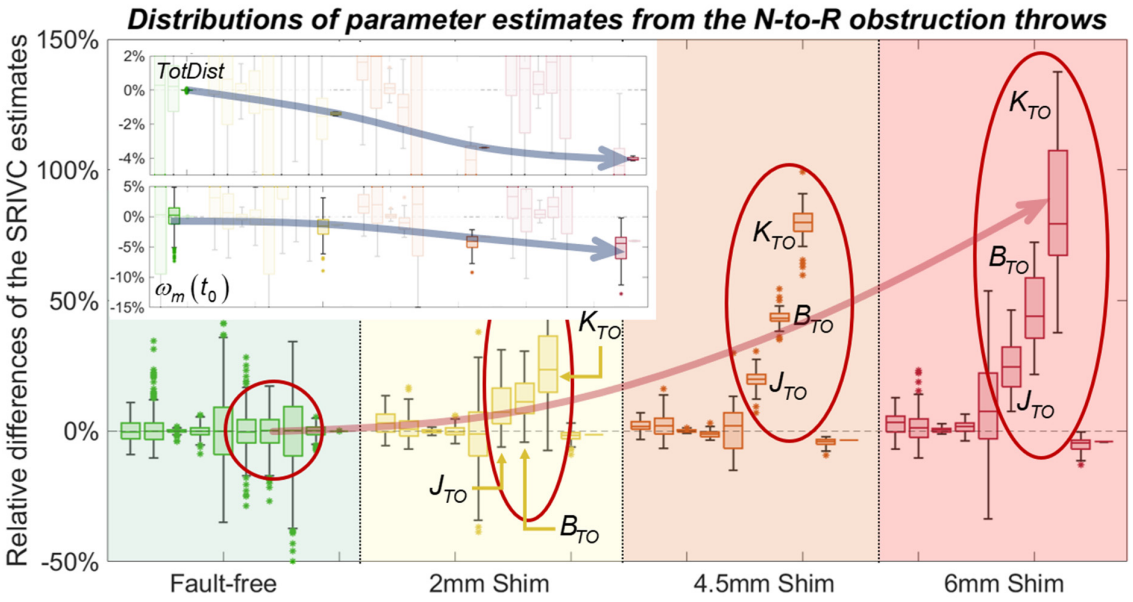


Figure 5-19 Estimation results of obstruction N-to-R throws simulated in three severity levels, with the zoom-in distributions of TotDist and $\omega_m(t_0)$

The boxplot distributions indicate that all mechanical parameters in the "move + lock" phase increase proportionally to the simulated fault severity. These parameter changes display another similarity of obstruction throws to the overdriving ones; even the estimates of $\omega_m(t_0)$ are also decreased when the faults get more severe.

The differences start to appear in the trend of changes in *TotDist*. This is intuitive because there is an obstruction in the opening gap, so the switch rails cannot be fully actuated compared to the fault-free throws. As the obstruction size gets larger, from 2 to 6 mm, the *TotDist* measured by the linear potentiometer displays a significant decreasing trend of up to 4%. Another difference lies in the magnitude of changes in all "move + lock" mechanical parameters. For example, at the "9-face" overdriving condition, the rise in B_{TO} and K_{TO} reach 100% and 200%, respectively. However, in the most severe obstruction condition (i.e. 6 mm shim), they are only 50% and 100%.

As a consequence, the reduction in total rail displacement has become the most dominant FDD feature for diagnosing obstruction. It is also accompanied by increases

in all mechanical parameters in the “move + lock” phase. Nevertheless, the magnitude of these changes is lower than those in the overdriving conditions.

5.3.2.2 OPP-Obstruction faults: estimation results

Once the metal shim is moved to the motor-side stock rail instead, N-to-R throws are now affected indirectly. As previously discussed, the effects in this condition are expected to be similar to those of unlock difficulty throws, but with the magnitude of changes being less significant. The comparison of measurements from every N-to-R OPP-obstruction throw from all severity levels is illustrated in Figure 5-20.

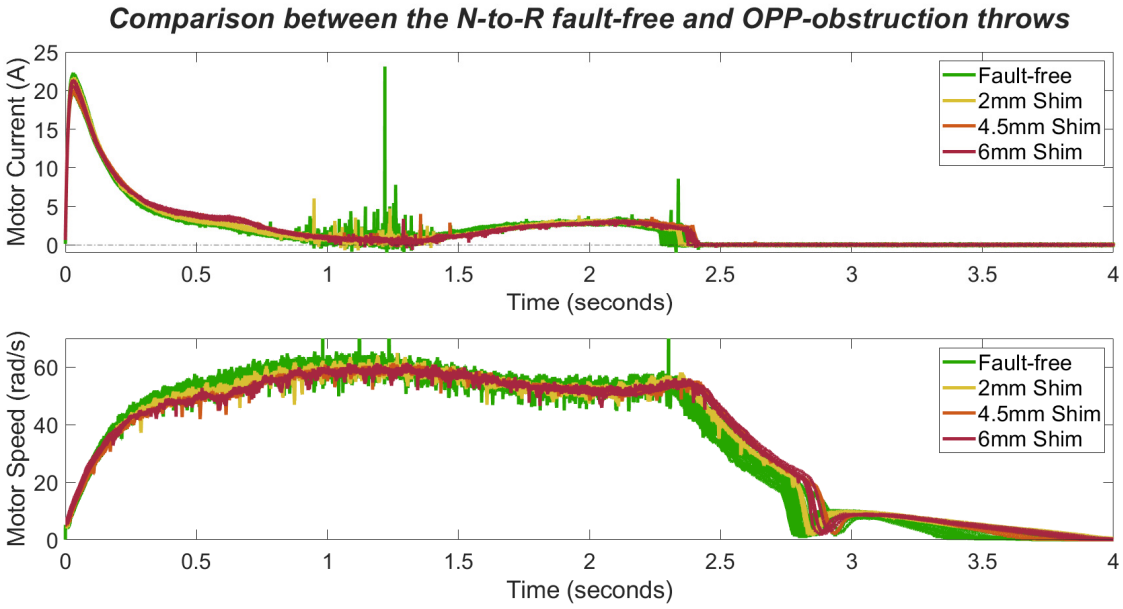


Figure 5-20 Comparison of motor current and motor speed measurements between fault-free and simulated OPP-obstruction N-to-R throws

The graphs reaffirm the hypothesis that the OPP-obstruction throws are similar to the unlock difficulty. During the unlocking phase, the motor current is slightly higher than in the fault-free condition, and the speed is also marginally lower. The differences start to appear in the estimation results, as shown in Figure 5-21.

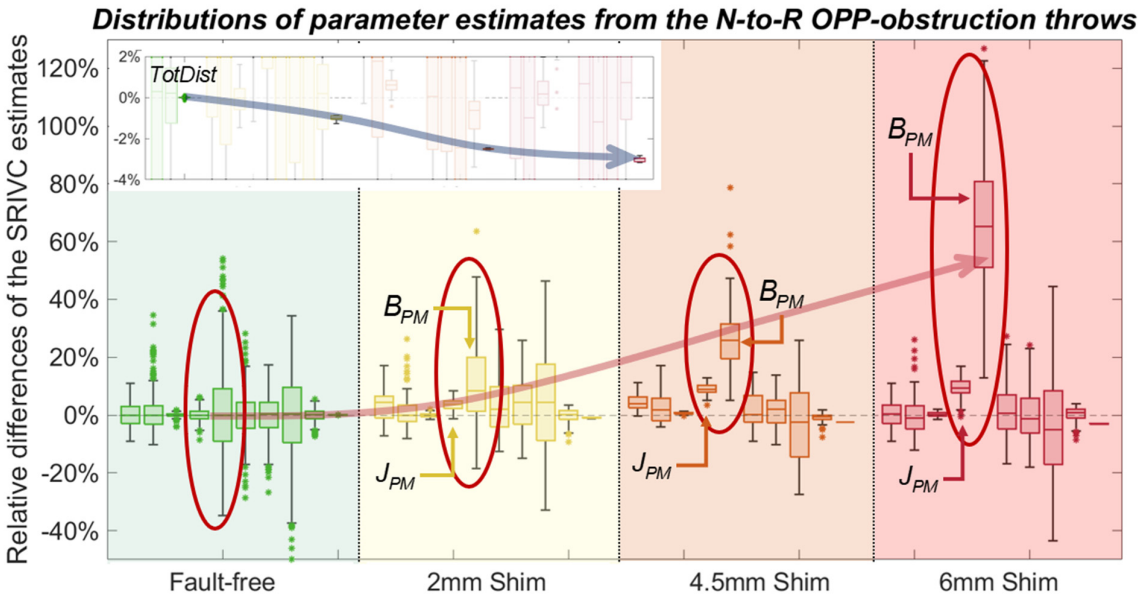


Figure 5-21 Estimation results of OPP-obstruction N-to-R throws simulated in three severity levels, with the zoom-in distributions of TotDist

The outstanding difference lies in the total rail displacement. If the switch rails cannot be fully actuated due to an obstruction during an R-to-N throw, they cannot be fully actuated in the other direction either. As a result, the decreasing trend of up to 4% in *TotDist* also appears in the N-to-R OPP-obstruction throws.

Accompanying the changes in *TotDist* are increases in mechanical parameters in the "inrush + unlock" phase. The estimates of B_{PM} are reiterated as less reliable FDD features because their distributions between severity levels are still highly overlapped. Although the level of changes in the estimates of J_{PM} is not as significant as in the unlock difficulty conditions, their distributions are still well separated between severity levels. These characteristics are due to the fact that a 6 mm shim produces a smaller overdriving effect compared to the "9-face" overdriving conditions. Finally, the OPP-Obstruction conditions also affect the *TotDist* considerably, which can be examined to decide the severity of obstruction faults in both directions as well.

In conclusion, the impacts from overdriving and obstruction faults, as seen by the DC motor, are indifferent. They both lead to the switch rails being hard-pressed during the locking phase. One is caused by lock nut maladjustment, leading to an excessive press between the switch rail and the stock rail. The latter makes it happen sooner by the insertion of a metal shim, and the interaction now occurs between the obstruction and the switch rail. As a result, *TotDist* has become the only parameter that can effectively differentiate the obstruction conditions from overdriving.

5.3.3 Underdrive fault simulations

Similar to overdrive fault occurrences, underdriving conditions also occur due to misalignments and out-of-adjustments. However, instead of the switch rails being hard-pressed against the stock rail, they are not even driven far enough to close the gap.

Underdrive faults are simulated by manually relocating the lock nuts, but in the opposite direction to that shown in Figure 5-11. As previously described in detail in Section 2.1.4.4, this maladjustment will extend the “lost-motion” distance but shorten the switch rail’s actuation stroke. If this stroke leaves a remaining gap of less than the locking mechanism’s allowance, it will only result in an underdrive fault. Otherwise, the lock notch will not be properly aligned with the lock dog, causing their engagement to be impossible and them to obstruct each other. The motor, however, will keep trying to pull the obstructed lock dog back to the locking position. Once the throw is not completed within the time limit, the power supply will be cut off, resulting in a failure.

Since the lock and the subsequent detection have failed, the signallers will be notified, and that track section will be shut down, waiting for corrective maintenance. This underdrive failure mechanism of an N-to-R throw is shown in Figure 5-22.

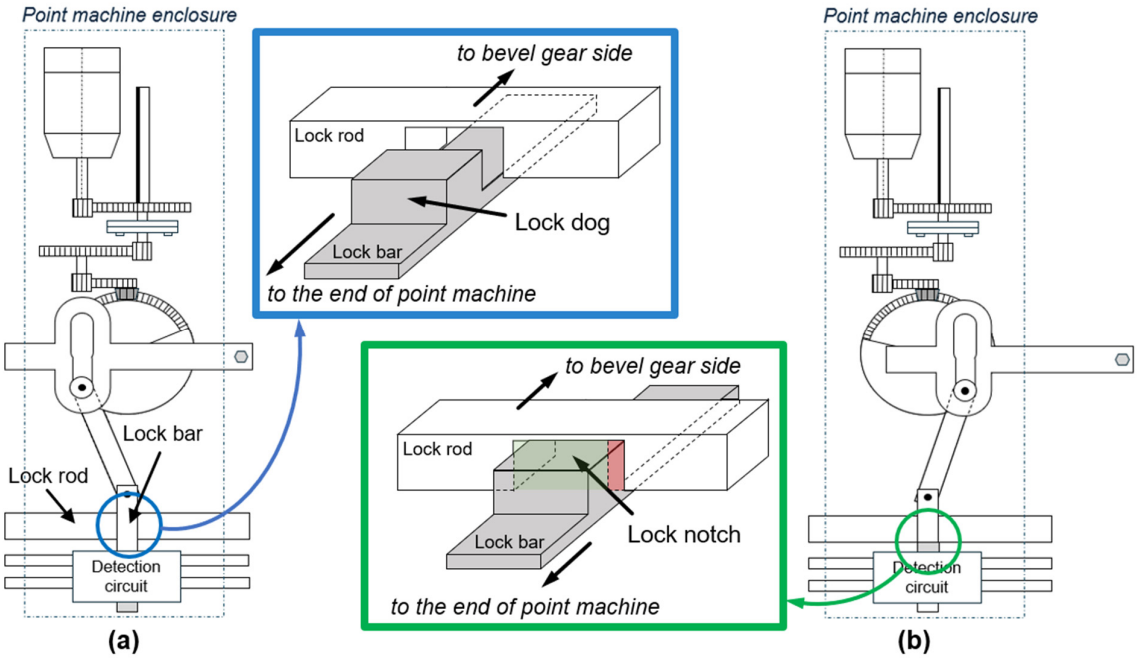


Figure 5-22 Details of HW55S locking mechanism using the coupling of the lock dog and the lock notch and their misalignment (shaded red area) caused by underdriving failures

To conduct a similar experiment compared to the previous two fault modes, the impacts on N-to-R throws when the far-side lock nuts are misplaced are also examined. This aims to determine whether there are any influences in the opposite direction that are not directly affected by the underdriving conditions. Therefore, these throws are defined as “OPP-underdrive”. Three severity levels are tested for both scenarios: the lock nuts will be turned to move away from the nominal position by “3-face” (half turn), “6-face” (one turn), and “9-face” (one and a half turns). Note that “10-face” is the first condition that causes underdriving failures. The following two sub-sections provide the parameter estimation results of N-to-R underdrive and OPP-underdrive throws.

5.3.3.1 Underdrive faults: estimation results

With the motor-side lock nuts being maladjusted, the motor current and speed measurements from N-to-R underdrive throws are compared in Figure 5-23.

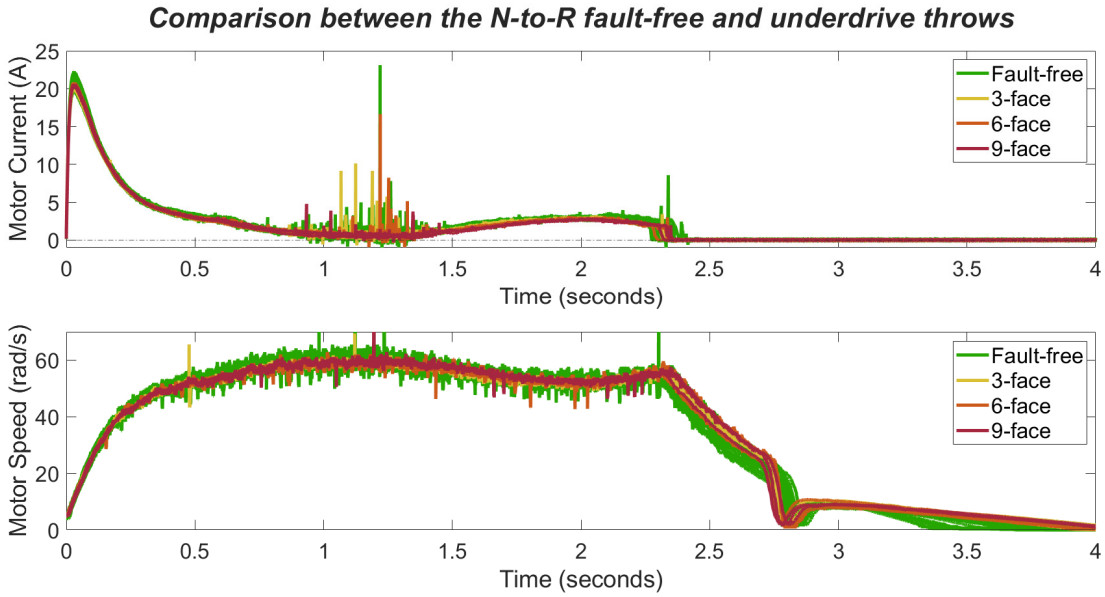


Figure 5-23 Comparison of motor current and motor speed measurements between fault-free and simulated underdrive N-to-R throws

These measurements demonstrate nearly perfect overlaps between the fault-free and underdrive throws at every severity level, with the only difference in their throw time. Since the underdriving mechanism shortens the actuation stroke, the underdrive throws are shown to be completed sooner than the fault-free ones. However, the throw time is still nearly identical at all severity levels, so more considerations are required to identify the fault severity more effectively. The estimation results of all simulated N-to-R underdrive throws are presented in Figure 5-24.

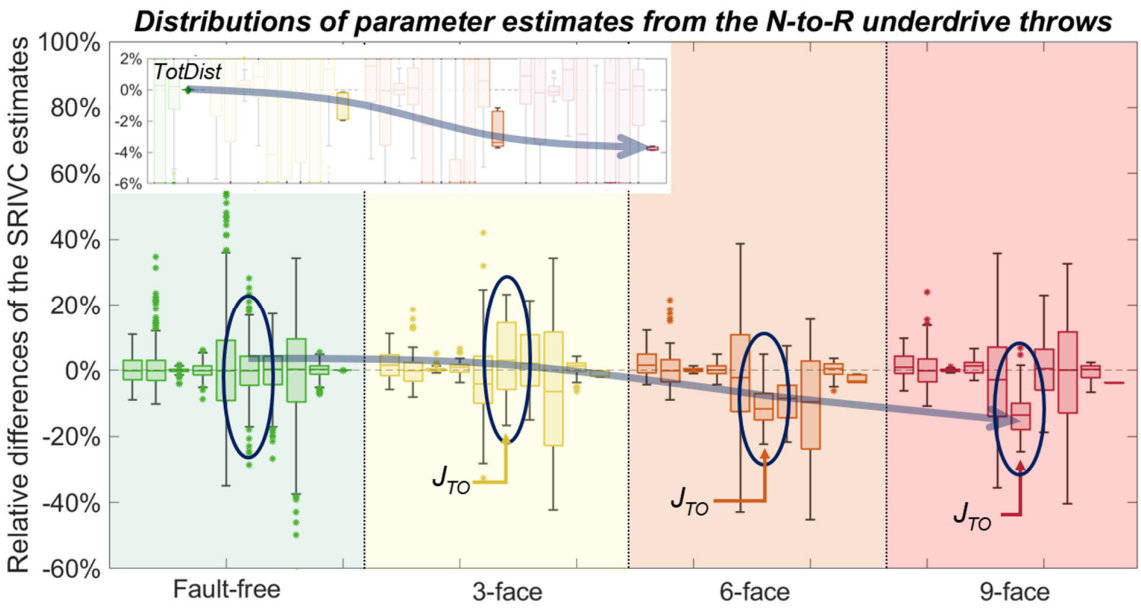


Figure 5-24 Estimation results of underdrive N-to-R throws simulated in three severity levels, with the zoom-in distributions of TotDist

First, the boxplots illustrate no significant effects resulting from the underdriving conditions. The only parameter that seems to be consistently affected is J_{TO}; it shows a decreasing trend of up to 20% at the most severe condition. However, due to such a small magnitude of changes and overlaps between severity levels, the decreases in J_{TO} would only be able to imply the occurrence of underdrive fault, not to indicate their criticality. Meanwhile, the estimates of B_{TO} change inconsistently, and the estimates of K_{TO} are too widely distributed to be used as effective FDD features.

With all parameter estimates being distributed around 0%, just like the fault-free conditions, the only parameter that possesses the most consistent declining trend with its magnitude corresponding to the severity level is the total rail displacement. As the lock nuts were maladjusted to shorten the actuation stroke, *TotDist* is shown to be lower than that of the fault-free throws. Moreover, as the lock nuts were moved even farther away from the drive lug, *TotDist* also decreased in proportion, making it the most relevant parameter for diagnosing underdriving conditions.

5.3.3.2 OPP-underdrive faults: estimation results

Once the motor-side lock nuts are adjusted back to their nominal position, the far-side lock nuts will be moved away from the drive lug instead. This experiment aims to observe the indirect effects of underdriving conditions when they are simulated in the opposite direction. All measurements obtained from the N-to-R OPP-underdrive throws from every severity level are illustrated in Figure 5-25.

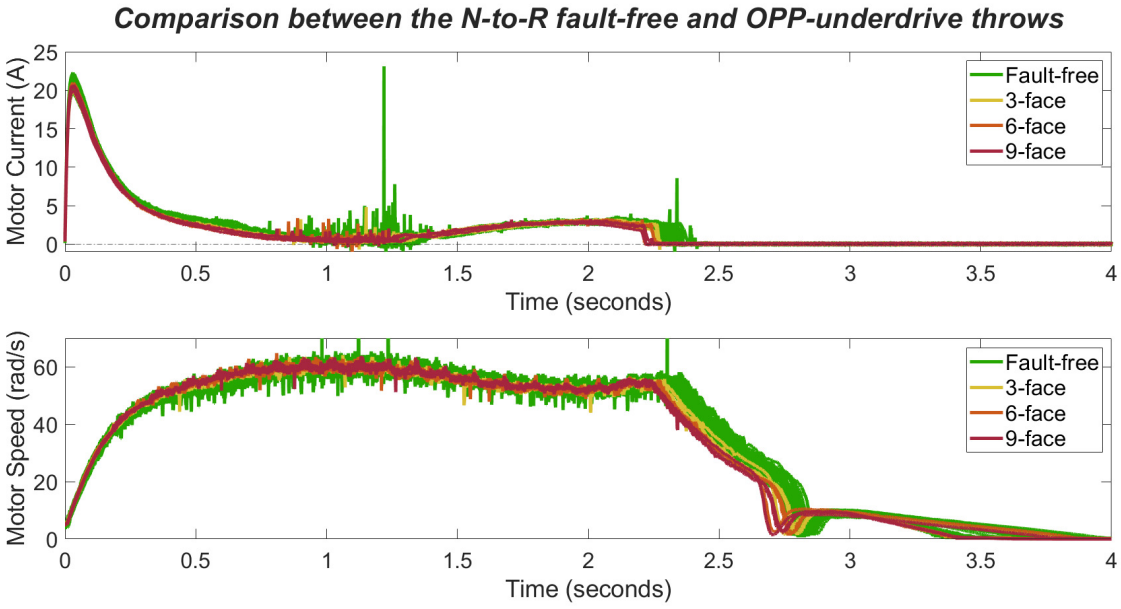


Figure 5-25 Comparison of motor current and motor speed measurements between fault-free and simulated OPP-underdrive N-to-R throws

Besides the throw time of the OPP-underdrive throws being shorter than that of the fault-free conditions, another subtle difference appears in the motor current during the unlocking phase. As the restoration force from the switch rail's stiffness becomes lower for underdriving conditions, the unlocking process of the switch rails becomes less restrictive as well, resulting in a lower motor current during this phase.

However, the differences in the measurements are indistinguishable between severity levels. More considerations must be conducted on the estimation results of the N-to-R OPP-underdrive throws, which are demonstrated in Figure 5-26.

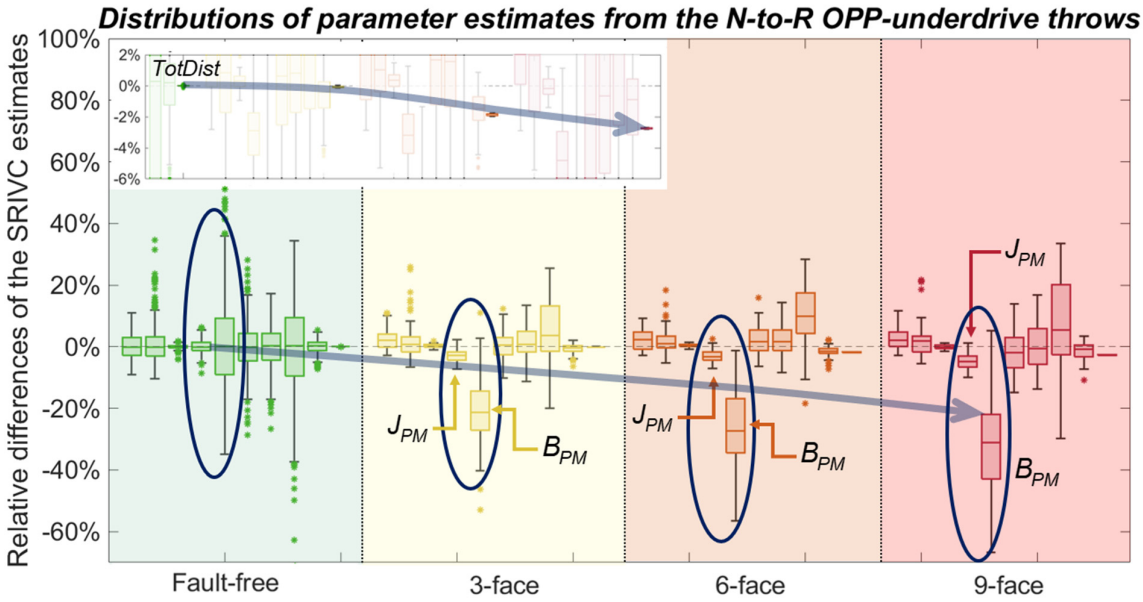


Figure 5-26 Estimation results of OPP-underdrive N-to-R throws simulated in three severity levels, with the zoom-in distributions of TotDist

Since the differences in the unlocking phase are more apparent than those in the locking one, the estimates of mechanical parameters in the “inrush + unlock” phase are also more prominent than those in the “move + lock” phase. In this scenario, both J_{PM} and B_{PM} are lower than those in fault-free conditions due to less actuation torque required to unlock the switch rails. However, the magnitude of changes in J_{PM} is subtle, with a maximum reduction of approximately 5% in the “9-face” condition. Also, the estimates of B_{PM} are still widely distributed. The reduction in both parameters, thus, would only be able to imply underdriving conditions occurring in the opposite direction.

The only turnout parameter that still proves to be the most relevant one capable of reflecting the severity level is the total rail displacement. Therefore, if the mechanical parameters in the “move + lock” phase decrease slightly with a considerable reduction in *TotDist*, these can imply the occurrence of underdrive faults. Meanwhile, if the estimates of J_{PM} and B_{PM} falls below their baseline, underdriving conditions may have occurred in the opposite direction of the considered throw.

5.3.4 Abnormal input voltage fault simulations

Normally, railway turnouts are energised by their associated LOC, which can be in the range of several kilometres apart. For HW55S turnouts, their nominal voltage of 110 V will be sent to the point machine from the power supply inside the lineside LOC. However, the voltage received at the DC motor's terminals could be at different levels. Although it can be directly measured by a voltage transducer fitted inside the point machine, its need for a separate power supply may not make it feasible. Figure 5-27 displays an overview of the drive circuit and where the voltages can be measured.

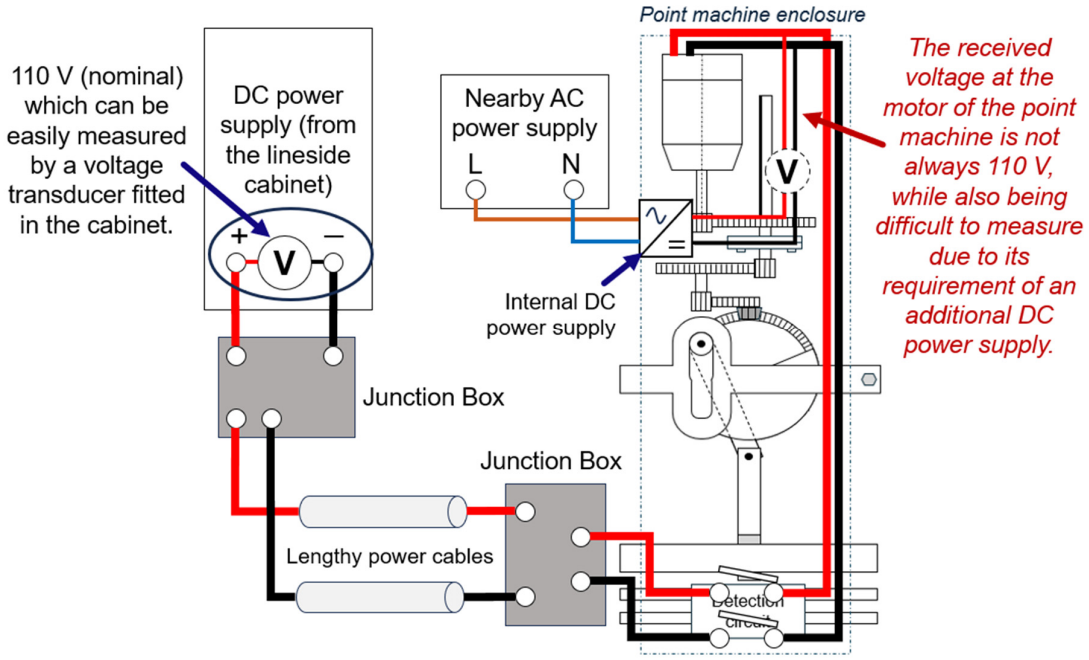


Figure 5-27 Drive circuit from the lineside LOC to the working turnout system displaying connections in between and where the voltages can be measured

In this thesis, the simulation of abnormal input voltage faults is conducted by adjusting the voltage level at the main power supply. Meanwhile, the input voltage data used in the estimation are assumed as a constant of 110 V.

This assumption represents the situation where the voltage transducer can be conveniently fitted inside the LOC and accessible, but the received voltage at the point

machine is completely unknown. Although a voltage transducer is included in the data acquisition system, as shown in Figure 5-5, its measurements are only used to verify the voltage level coming out from the power supply.

Four simulation scenarios, 120 V, 110 V (fault-free), 100 V, and 90 V, represent the four severity levels. The motor current and speed measurements from this fault simulation in N-to-R throws at all severity levels are compared in Figure 5-28.

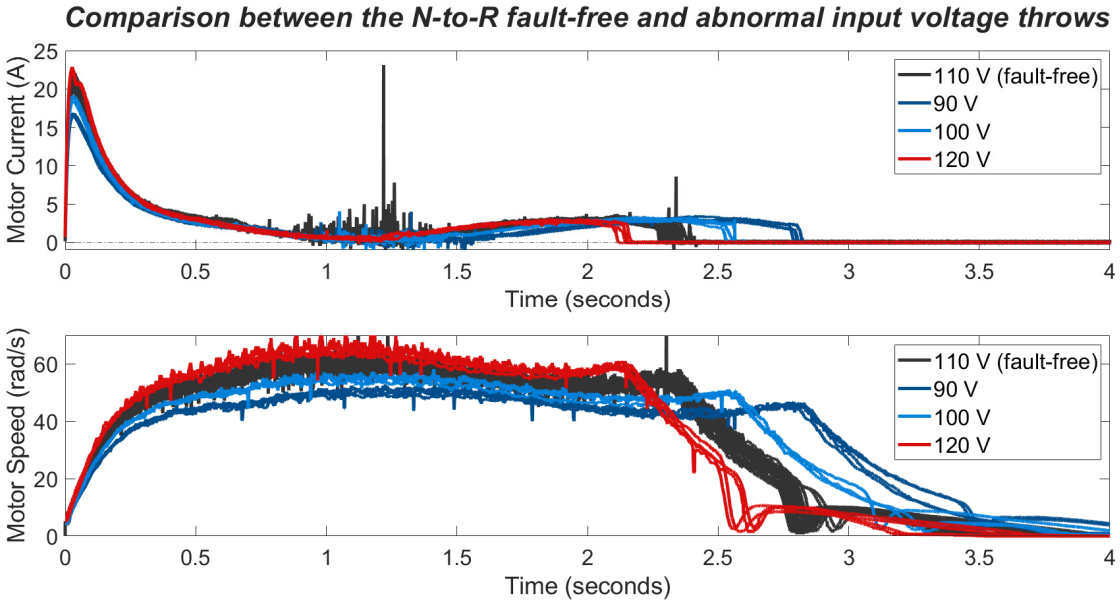


Figure 5-28 Comparison of motor current and motor speed measurements between fault-free and simulated abnormal input voltage N-to-R throws

The plots display clear distinctions between voltage levels. When the voltage is lower than the nominal, the motor speed reduces considerably causing the throws to be completed much later than that in the fault-free condition. If the received voltage is much lower such that the throw time is longer than the timer relay setting, the throws will be intermediately interrupted, resulting in turnout failures. In contrast, when the voltage is higher than 110 V, the motor speed increases proportionally, and the throws are completed sooner by approximately 0.2 seconds in the 120 V scenarios.

Another important feature in the measurements is the peak inrush current. In cases where the voltage is lower than 110 V, the point machine can only generate an actuation torque of less than the nominal. Therefore, the peak current where the input voltage is either 100 V or 90 V is lower than that of the fault-free conditions. On the other hand, when the input voltage is 120 V, the peak current is the highest (compared to all other cases). This leads to higher torque, faster motor speed, and shorter throw time, which potentially generates more heat inside the point machine enclosure.

Once the outstanding features from the measurement comparison have been considered, the parameter estimation results of the simulated abnormal input voltage N-to-R throws are now illustrated in Figure 5-29.

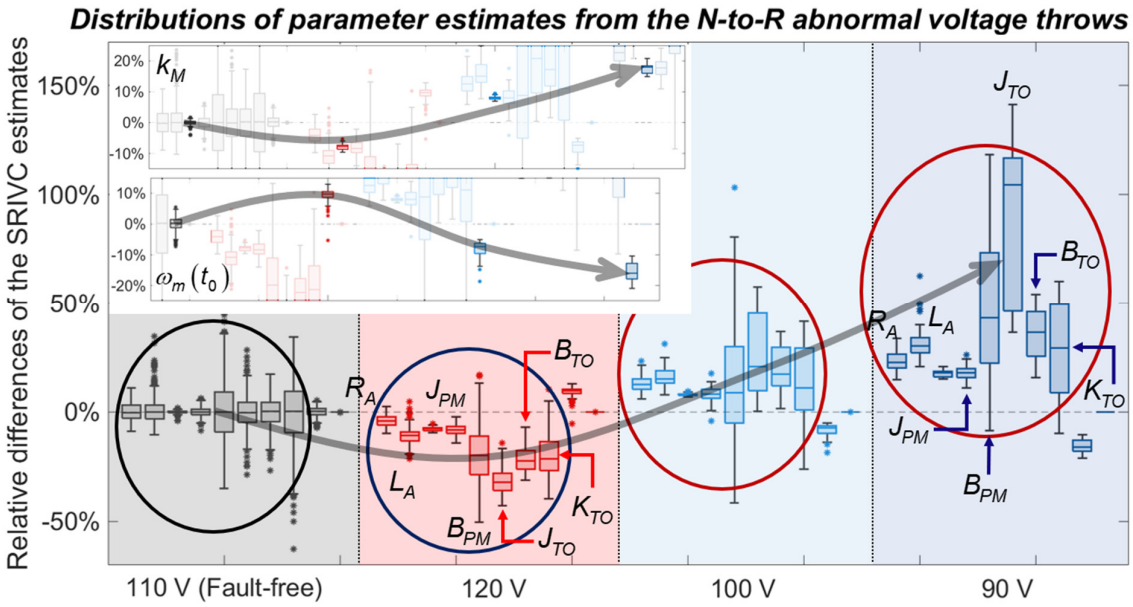


Figure 5-29 Estimation results of abnormal input voltage N-to-R throws simulated in three severity levels, with the zoom-in distributions of k_M and $\omega_m(t_0)$ estimates

Since this fault affects the electrical characteristics of the HW55S turnout, both estimates of R_A and L_A are clearly impacted. As previously stated, when the voltage is lower than the nominal, the peak inrush current is also reduced. As a nominal voltage of 110 V is still assumed in the estimation, it will produce higher estimates of both R_A

and L_A instead, and vice versa. However, their distributions between severity levels are fairly overlapped, making them less effective in identifying the fault severity. One parameter that is exceptional in terms of changing in accordance with the severity levels and possessing no overlap in distributions at all is the motor constant.

In Chapter 4, the changes in motor constant were shown to influence all of the mechanical parameter estimates, as it is assumed that a permanent magnet DC motor is utilised, which has a relationship of $k_M = k_T$. This assumption actually holds for the HW55S turnouts. Thus, the qualitative changes in all mechanical parameters in both the “inrush + unlock” and “move + lock” phases have been demonstrated to coincide with the direction of changes in the k_M, whereas the distributions of J_{PM}, J_{TO}, and B_{TO} estimates are better separated between severity levels compared to the others.

The final important feature lies in the estimates of the initial speed entering the moving phase, i.e. $\omega_m(t_0)$. Due to the dynamics of DC motors, when they are supplied with a higher voltage, their shaft will rotate faster. This relation is reflected in $\omega_m(t_0)$ and *TotDist* parameters as well, as their estimates change correspondingly with the actual input voltage level. However, the effects on *TotDist* are less prevalent.

Therefore, to detect abnormal input voltage received at the point machine, every parameter estimate will be different compared to its baseline. Changes in all parameter estimates will be inversely proportional compared to the actual received voltage level. However, the estimates of $\omega_m(t_0)$ will be changed proportionally instead. The degree of changes in k_M and $\omega_m(t_0)$ are also shown to be indicative of the severity level.

5.3.5 Improper sliding surface fault simulations

The final fault mode to be investigated is the improper sliding surface, which has already been studied with the turnout simulation model in Chapter 4. In the simulation

study, the coefficient of friction was increased to impose more resistance against the switch rail's actuation. Since the available HW55S turnout system uses slide chairs to ease the actuation effort, three conditions will be applied to their surface as follows:

- Fault-free: the experiments were conducted without any alteration;
- Dry slide chairs: the existing lubrication layer that was applied two months prior to the data collection campaign was first wiped out from every slide chair. Then, a solution was applied to make sure that they were completely dry and clean and had as little dust as applicably possible;
- Greased slide chairs: once measurements from all "dry slide chair" throws were obtained, every slide chair was newly greased and distributed evenly. After that, several throws were completed first to ensure that the grease went under the switch rails and made contact with the slide chairs.

Photos of slide chairs from all three conditions are shown in Figure 5-30 to 5-32.



Figure 5-30 Samples of the fault-free slide chair condition

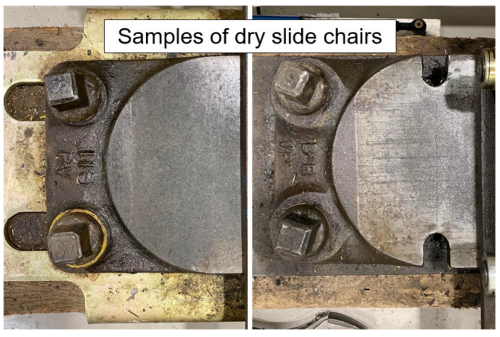


Figure 5-31 Samples of the dry slide chair condition



Figure 5-32 Samples of the greased slide chair condition

The comparison between measurements of simulated N-to-R improper sliding surface throws from all three conditions is displayed in Figure 5-33.

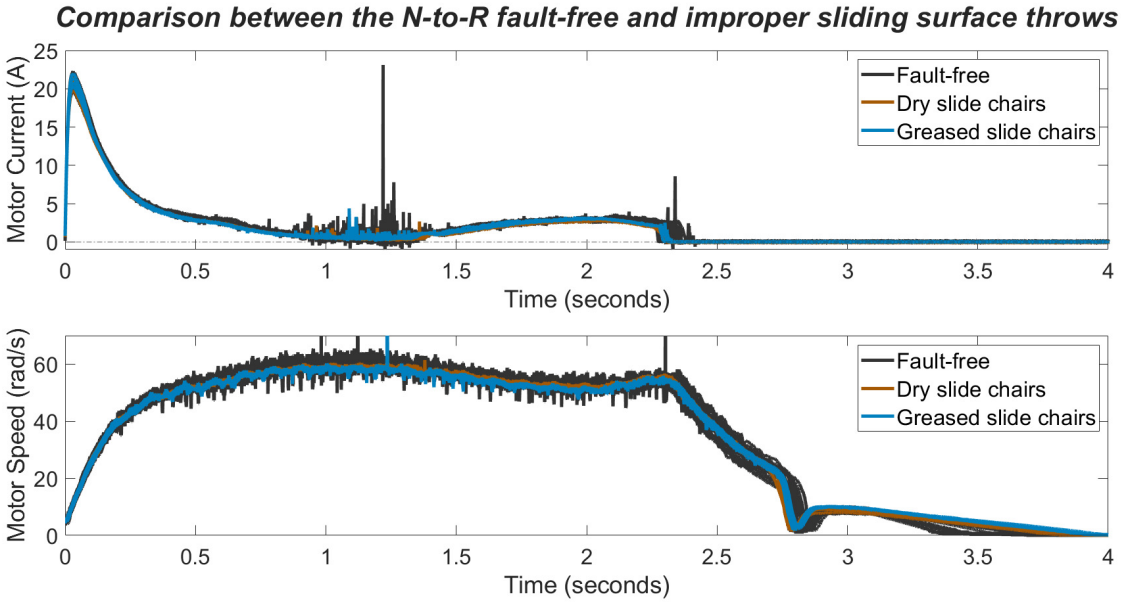


Figure 5-33 Comparison of motor current and motor speed measurements between fault-free and simulated improper sliding surface N-to-R throws

Regardless of the condition of the slide chair's surface, the measurements from all three conditions are nearly overlapped on each other. This characteristic is actually counterintuitive as when the slide chairs are dry or contaminated, the overall friction during the moving phase of the operation is expected to increase. On the contrary, when they are recently well-lubricated, system friction should be reduced.

To verify these counterintuitions, the estimation results extracted from all three slide chair's surface conditions are illustrated in Figure 5-34.

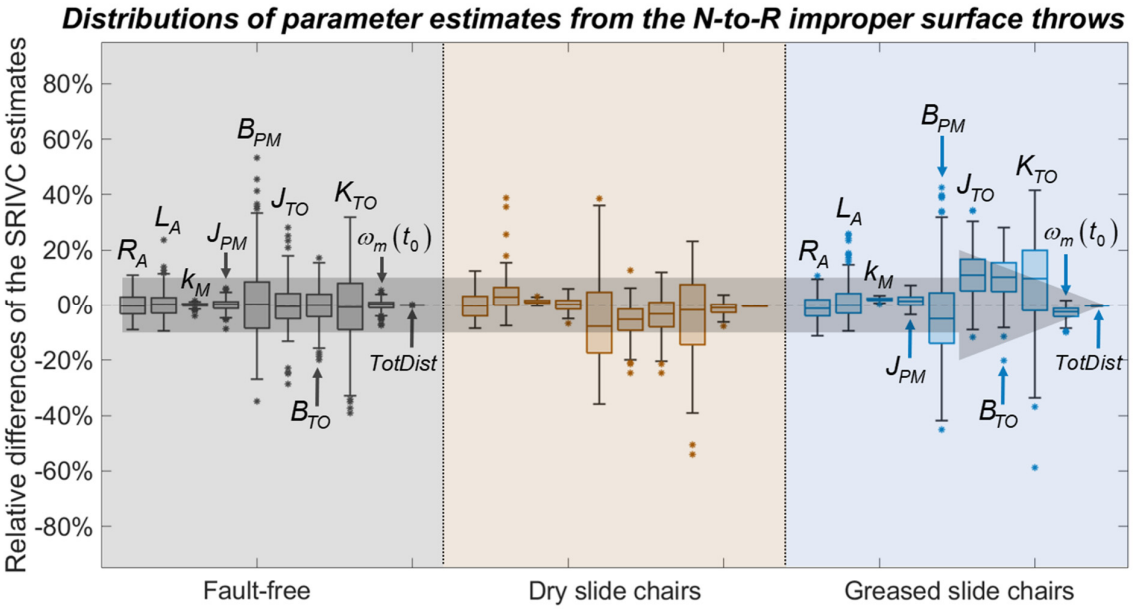


Figure 5-34 Estimation results of improper sliding surface N-to-R throws simulated in three slide chairs' surface conditions

The distributions of all turnout parameters reaffirm that this fault simulation does not significantly affect the HW55S turnout. Although the estimates of J_{TO} and B_{TO} in Figure 5-34 may indicate a sign of a changing trend; their magnitude is subtle and their distributions are heavily overlapped, causing them not to be relevant FDD features.

Nevertheless, scenarios where the effects of improper sliding surface faults are unobservable have been previously mentioned by Márquez, Weston, and Roberts (2007). One of the reasons was that "the switch rails may not make contact with the slide chairs". Thus, the interface between the switch rails and the slide chair's surface is now being closely inspected, with the finding of a small gap shown in Figure 5-35.

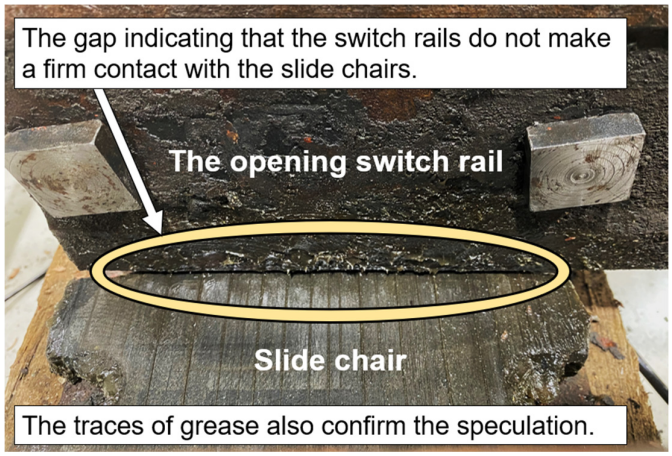


Figure 5-35 Inspection of the interface between the switch rails and the slide chairs

As a result, it can be concluded that if the switch rails of the turnout system do not make firm contact with the slide chairs, as displayed in Figure 5-35, the proposed parameter estimation will not be able to detect improper sliding surface faults. This is due to the fact that the slide chairs do not have a significant impact on the actuation, causing the changes in the parameter estimates negligible and inconsistent.

The experiments of all fault modes considered in this thesis are now completed. The following and final section will summarise all the findings from each fault simulation and generate the essential parameter change “knowledge-base” table.

5.4 Parameter change “knowledge-base” table

This final section aims to summarise all findings from previous fault simulations to generate a parameter change “knowledge-base” table. Table 5-5 summarises the information on which turnout parameter estimates are related to each fault mode that will later facilitate the FDD algorithm development in the following chapter.

Table 5-5 Parameter change knowledge-base table summarising the correlation between the considered fault modes and the trend of changes of every turnout parameter

Fault modes	**Turnout parameters**									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Overdrive						↑	↑↑	↑↑	↓	↑↑
Unlock Difficulty				↑↑	↑					↑↑
Obstruction						↑	↑↑	↑↑	↓	↓↓
OPP-obstruction				↑↑	↑					↓↓
Underdrive						↓				↓↓
OPP-underdrive				↓	↓					↓↓
Overvoltage	↓	↓	↓↓	↓↓	↓	↓	↓↓	↓	↑↑	↑
Undervoltage	↑	↑	↑↑	↑↑	↑	↑	↑↑	↑	↓↓	↓
Improper sliding surface										

NOTE1: Up-arrows represent a direct proportional correlation of the parameter values to the severity levels, while down-arrows represent an inversely proportional trend.

NOTE2: The columns J_{PM} and B_{PM} are highlighted in blue to indicate that they are only considered during the "inrush + unlock" phase, while the columns J_{TO}, B_{TO}, K_{TO}, and $\omega_m(t_0)$ are highlighted in orange to indicate that they are only considered during the "move + lock".

In the table, the single arrow sign only indicates which parameter is affected by the fault simulation, whereas the double arrows indicate that such a parameter is also a dominant FDD feature that can reflect the fault severity as well. If a cell is shaded with a grey colour and blank, that parameter is not affected by the simulated fault.

For all three mechanical fault modes that the proposed estimation methodology could detect their influences, only the mechanical turnout parameter estimates are different from their baseline. Depending on the considered throw direction, the effects may appear exclusively during the "inrush + unlock" or "move + lock" phase. Faults that affect the rail actuation stroke also translated its effects to the *TotDist* parameter.

Meanwhile, the abnormal input voltage fault simulation has been shown to affect every parameter. The impacts on the electrical parameters are intuitive, whereas those on the mechanical ones are indirectly affected due to the relationship of $k_M = k_T$.

Since Table 5-5 has been generated, the features to be emphasised during the FDD algorithm development are now determined. Compared to features often used in other data-driven methodologies, all estimated parameters in Table 5-5 are physically interpretable, related to the physical characteristics of the turnout, easy to understand, and capable of diagnosing different fault modes at different severity levels.

5.5 Conclusions

In this chapter, the proposed parameter estimation methodology was applied to a real HW55S turnout system. The initial results were unsuccessful due to the non-linear dynamics of the real system, which was different compared to the assumptions used in simulation studies during Chapters 3 and 4. It was found that the motor of the HW55S point machine experiences two distinct load conditions. As a consequence, a segmentation algorithm was developed to divide each operation into two linear phases. This process also required the transfer function models to be modified, specifically due to the inclusion of non-zero initial conditions. Having done this, the modified estimation algorithm was re-applied to the HW55S measurements, which was now able to provide exceptional data-fitting performance and ten reliable turnout parameter estimates.

Then, the five fault modes previously described in Chapter 2 were simulated on the HW55S turnout system. The parameter estimates obtained from these faults were compared to the baseline of each parameter (obtained from a number of known fault-free operations) to calculate the differences and facilitate parameter change detection.

Finally, from all fault simulations, a knowledge-base table has been formulated at the end of this chapter. It contains important information on which parameters are the relevant FDD features when diagnosing a specific fault mode. Such a knowledge base will be essential for the FDD algorithm development in the following chapter.

All in all, this chapter has demonstrated that the proposed parameter estimation methodology and integration of a segmentation algorithm can estimate the physically interpretable parameters of a real HW55S turnout system. Patterns in their differences, when the turnout system was in fault-free and faulty conditions, can also be mapped to the corresponding fault modes, offering potential fault diagnosis capability. This contributes to the creation of the knowledge-base table that will be used during the FDD algorithm development. All these results presented the first application of such a model-based approach to a real turnout system, which is the main novelty of this thesis.

CHAPTER 6 DEVELOPMENT OF FAULT DETECTION AND DIAGNOSIS ALGORITHM

Having concluded, in Chapter 5, that it is possible to monitor the current health state of a turnout based on estimates of its physical parameters, evaluation of changes and patterns in these parameters is mandatory and still requires expertise from the maintenance staff. Thus, for these FDD features to be useful for maintenance planning, they must be further evaluated to produce diagnosis information automatically.

This chapter aims to develop such an FDD algorithm that can also ensure the thorough interpretability and traceability of the proposed methodology, entirely from the features to the diagnostic results. Such characteristics would allow users to self-validate the results based on their personal experience, enabling interactivity and improving trust towards using this automatic FDD methodology. So, these parameter estimates will be evaluated by a rule-based classification, which must be adequately transparent to enable noncomplex reconfiguration when the methodology is applied to other turnout systems as well. The main contributions of this chapter are:

1) The rule-based FDD classification will be developed and tested;

2) Then, the transferability of the proposed FDD methodology will be assessed by applying it to a different HW2000 turnout system.

6.1 Rule-based FDD classification algorithm

The feature extraction process presented in Chapter 5 can effectively estimate interpretable parameters from the raw measurements. To maintain such interpretability thoroughly for the entire methodology, the following feature evaluation process must also be interpretable. As highlighted in Chapter 2, commonly used machine learning

and deep learning-based classifiers are not suitable for this purpose due to their lack of explainability. Therefore, an intuitive rule-based FDD classification was chosen.

Maintenance experts are normally excellent at detecting significant changes in these FDD features and extracting their patterns to deduce diagnostic outcomes. With the intention to replicate such approximated human reasoning, the concept of fuzzy logic will also be integrated into the rule-based FDD classification development in this thesis (Wang and Li, 2024). The following section introduces the fuzzy logic principle and elaborates on its application to the proposed FDD methodology.

6.2 Fuzzy logic rule-based classification

In this section, the application of fuzzy logic in order to classify the interpretable parameter estimates into their most likely associated fault mode will be described. First, the fuzzy logic and linguistic variables will be briefly introduced, and their distinct benefits compared to traditional Boolean logic will be highlighted. Then, procedures for employing fuzzy logic for FDD purposes will be subsequently elaborated, including fuzzification, creation of the fuzzy inference system, and defuzzification.

6.2.1 Introduction of fuzzy logic

Classical Boolean logic can only output either ‘0’ or ‘1’. So, when it is applied for fault detection, these logical expressions are translated into “completely fault-free” and “faulty”, not allowing intermediate decisions. Such limitation is not adequate for FDD applications in turnout systems because their deterioration state from being fault-free to being faulty is also imperative for scheduling effective maintenance.

Fuzzy logic has been demonstrated to be a viable solution in previous studies as it is capable of modelling logical reasoning that can produce non-binary outcomes (Zadeh, 1965; Jafarian and Rezvani, 2012; Asadi and Shahrabi, 2016).

In contrast to Boolean logic, a set of membership values (which can be any number between 0 and 1) is assigned to each input variable to describe the degree of membership of that input to all linguistic fuzzy sets.

The procedure above is commonly known as “fuzzification”, as it “fuzzifies” the input variables into linguistic terms that vaguely describe their property. For instance, assume that there is a relative difference of the armature resistance estimate (i.e. R_A) at +20%, which will be fuzzified into “Low”, “Normal”, and “High” sets. The membership value of “High” would be greater than that of “Normal”, and since the actual resistance estimate is higher than the baseline, the membership value of “Low” would be zero.

The fuzzification process can be deployed to the fault detection results as well. Due to the logical expressions of fuzzy logic having the property of no longer being binary, deterioration states between “completely fault-free” and “faulty” can be inferred from the membership values of these two fuzzy sets. This interpretation of the output fuzzy variables to the final diagnostic result is called “defuzzification”, which should be as comprehensible as possible since it will directly interface with the users.

Finally, the mapping between the input fuzzy linguistic variables and the fuzzy outputs will be conducted in a fuzzy inference system (FIS). It consists of multiple fuzzy if-then rules where the fuzzy inputs are linked together by operators analogous to classical Boolean algebra, such as NOT, AND, and OR (Zadeh, 1975).

In this thesis, the fuzzy rules inside the FIS are derived from the knowledge base table, previously summarised in Table 5-5, since it already completely informs the relationships between the parameter qualitative changes and their corresponding fault modes. According to these descriptions of each element, a high-level workflow of such a fuzzy logic-based FDD classification algorithm is shown in Figure 6-1.

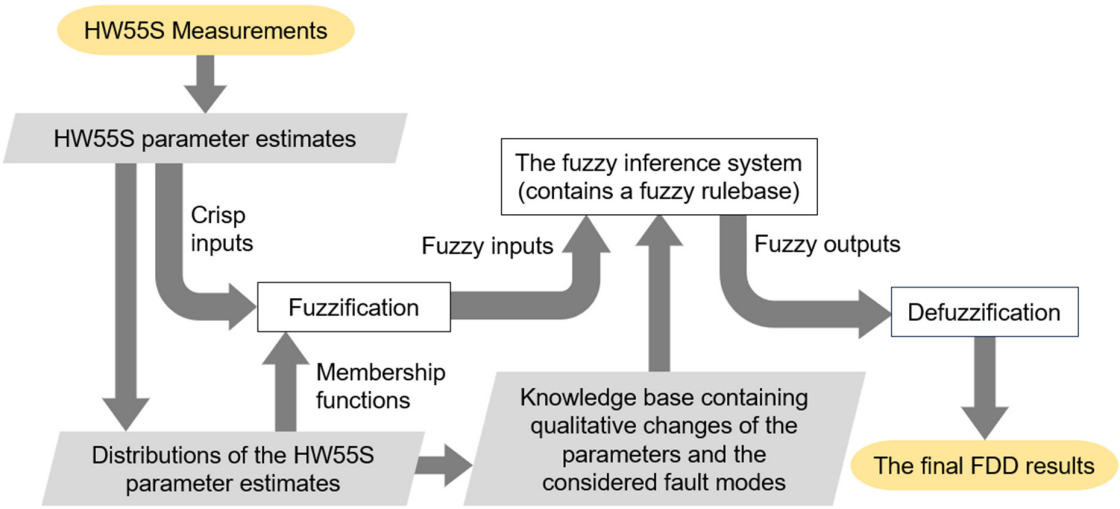


Figure 6-1 High-level workflow diagram of the proposed fuzzy logic-based FDD algorithm

The following sub-sections elaborate on how each element of the algorithm is developed for the proposed FDD methodology, while the main characteristics that are intended to be preserved are thorough interpretability and traceability.

6.2.2 Fuzzification of turnout parameter estimates

The first procedure is to transform all numerical turnout parameter estimates into their respective linguistic variables. The fuzzy sets that will be utilised to describe them are the magnitude of relative differences compared to their baseline. To simplify and preserve the comprehensibility of the algorithm, the number of fuzzy sets in this thesis is arbitrarily selected as three: “Low”, “Normal”, and “High”. These represent when the parameter estimates are less than the baseline, in the same range as the baseline, and more than the baseline, respectively.

Apart from the number of fuzzy sets, there are many membership functions to select from as well, such as trapezoidal, triangular, gaussian, and sigmoid (Antong, Dixon, and Ward, 2017; Narges et al., 2022; Mathworks®, 2023). Both the amount and related hyperparameters of the membership functions can also be optimally adjusted to maximise the algorithm’s performance. However, to maintain the traceability of the

proposed fuzzy logic-based FDD algorithm, the membership functions employed in this thesis are neither created arbitrarily nor based on the optimisation results. They will originate from **the statistical characteristics of the HW55S parameter estimates**.

As presented in Chapter 5, just under 2,500 estimation results were obtained from the HW55S measurements operated in all N-to-R conditions. It was found that when a particular turnout parameter is considered, the number of operations that does not affect the parameter significantly is higher than the number that does.

To elaborate on this statement, Table 5-5 indicates that the turnout's damping parameter (i.e. B_{TO}) is not impacted in fault-free, unlock difficulty, underdrive, OPP-underdrive, OPP-obstruction, and improper sliding surface scenarios. Considering only N-to-R operations, these were collectively 1,640 throws, which contributed more than 50% of the entire N-to-R experiment. As a consequence, the statistical characteristics of B_{TO} estimates from these throws are fitted to represent the entire population. Figure 6-2 illustrates the distribution of such unaffected B_{TO} estimates from all 1,640 N-to-R throws using histograms, an approximated distribution curve, and a boxplot.

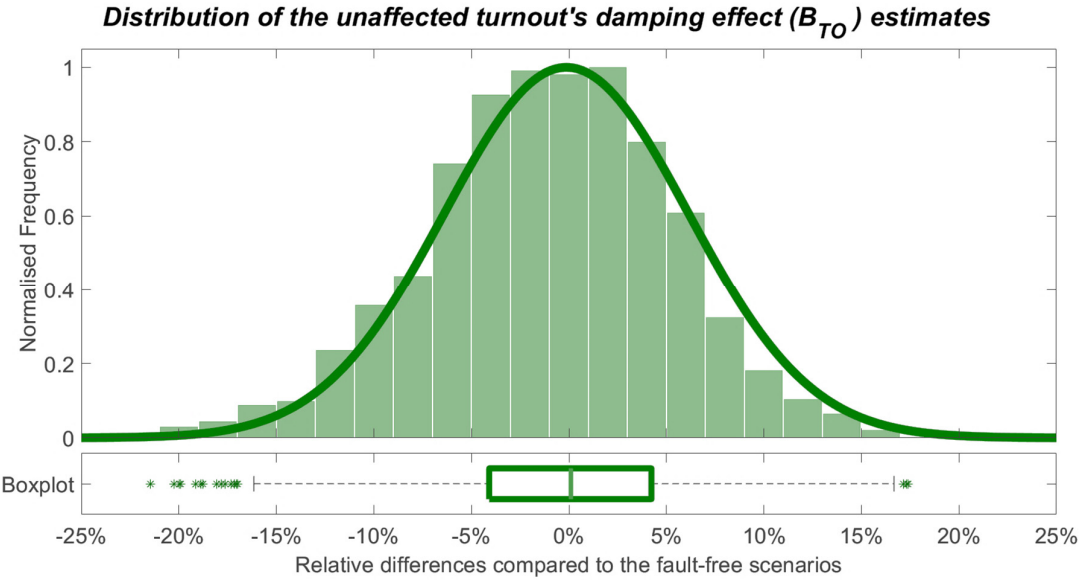


Figure 6-2 Distribution of unaffected B_{TO} parameter estimates obtained from N-to-R throws presented in histogram, approximated distribution curve, and boxplot formats

The figure displays a near balanced Gaussian distribution, where its centre is in the proximity of 0% difference. This near symmetrical Gaussian distribution trend not only holds for B_{TO} estimates, but for all other turnout parameters that are not affected during fault simulations as well, as shown in Figure 6-3. Such statistical characteristics will be the foundation for the following fuzzy set formulation.

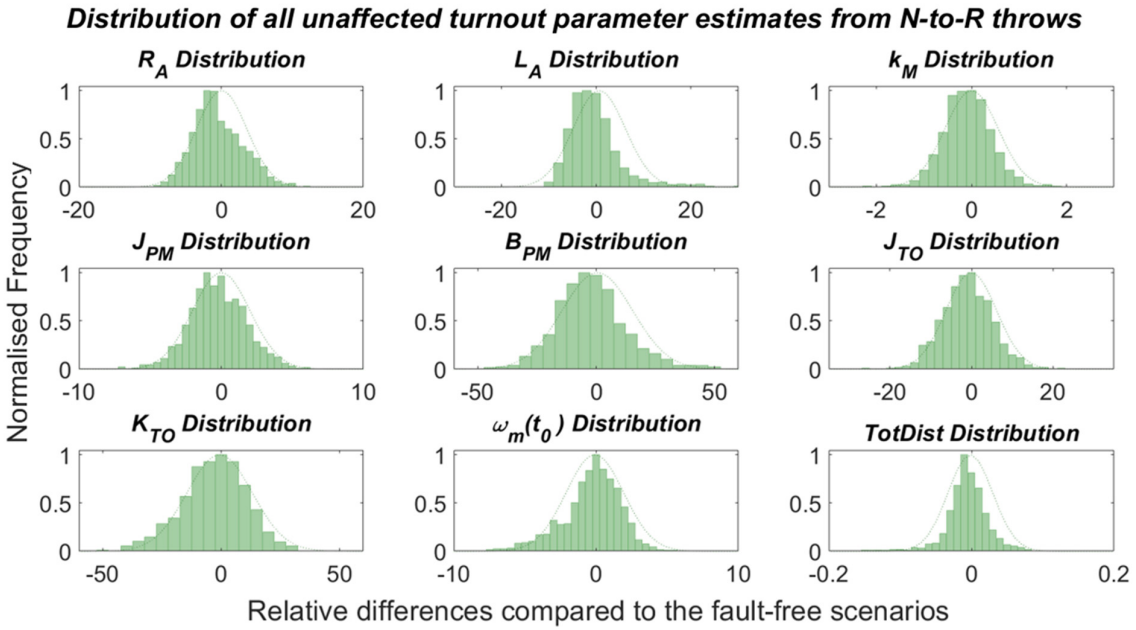


Figure 6-3 Distribution of other unaffected parameter estimates obtained from N-to-R throws

For example, the near balanced Gaussian distribution indicating the likelihood of the relative difference values of the B_{TO} estimates (shown in Figure 6-2) are now converted into the "Normal" fuzzy set of the input fuzzy variable "B_{TO}".

To cover the entire universe of discourse of B_{TO}, two other fuzzy sets, "Low" and "High", will be respectively placed on the left- and right-hand side of the "Normal" fuzzy set, based on the principle that the sum of all membership values should equal unity (Guillaume and Charnomordic, 2011). The resulting membership functions for the B_{TO} estimates are now formulated as demonstrated in Figure 6-4. For all other parameters, their fuzzy sets will be formulated by the same procedure.

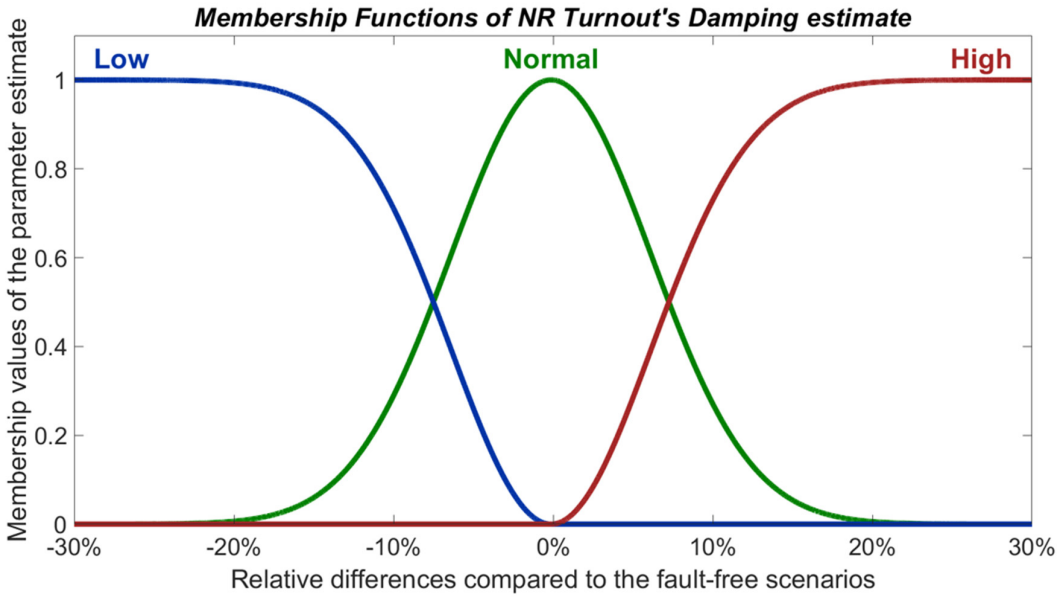


Figure 6-4 Example of the membership functions formulation for the B_{TO} estimates

The result of this sub-section is the fuzzification process for the input variables where the originality of all fuzzy sets is transparent, preserving the interpretability and traceability of the methodology in the process. The next sub-section describes how to integrate the input fuzzification results into formulating the fault detection fuzzy rule.

6.2.3 Formulation of fault detection rule in the fuzzy inference system

In this sub-section, the development of the fuzzy inference system (FIS) is initiated, which holds a collection of fuzzy rules used to produce the FDD results by analysing the input fuzzy variables. The first rule to be formulated is for fault detection purposes, which attempts to provide information on the level of turnout degradation.

The turnout will be regarded as completely faulty (with the membership value of "fault-free" being zero) if it is severely affected by any of the eight fault conditions listed in Table 5-5. As a consequence, the fault detection rule will be formulated based on **the effects of all eight fault conditions when they are at the highest severity level**, including "9-face overdrive", "9-face unlock difficulty", "6mm obstruction", "6mm OPP-obstruction", "9-face underdrive", "9-face OPP-underdrive", "120 V overvoltage", and

"90 V undervoltage". Suppose the estimation result of a testing throw does not relate to any of these conditions. In that case, the turnout will be regarded as fault-free.

First, considering faults that reduce the switch rail actuation stroke, including "9-face underdrive", "9-face OPP-underdrive", "6mm obstruction", and "6mm OPP-obstruction", these conditions produce a *TotDist* that is considerably lower than the fault-free scenarios. This would result in **the membership values of *TotDist* related to the "Low" fuzzy set (*TotDist_*μLow) being close to unity**, which is reaffirmed by the distributions of membership values from all three fuzzy sets (i.e. "*TotDist_*μLow", "*TotDist_*μNorm", and "*TotDist_*μHigh") of all four fault conditions in Figure 6-5. Due to such identical effects on *TotDist* from all these faults, the fuzzy rule that indicates the occurrence of either one of them can be expressed as follows:

If *TotDist* is Low
Then State is severe Underdrive OR severe OPP-underdrive
OR severe Obstruction
OR severe OPP-obstruction (6-1)

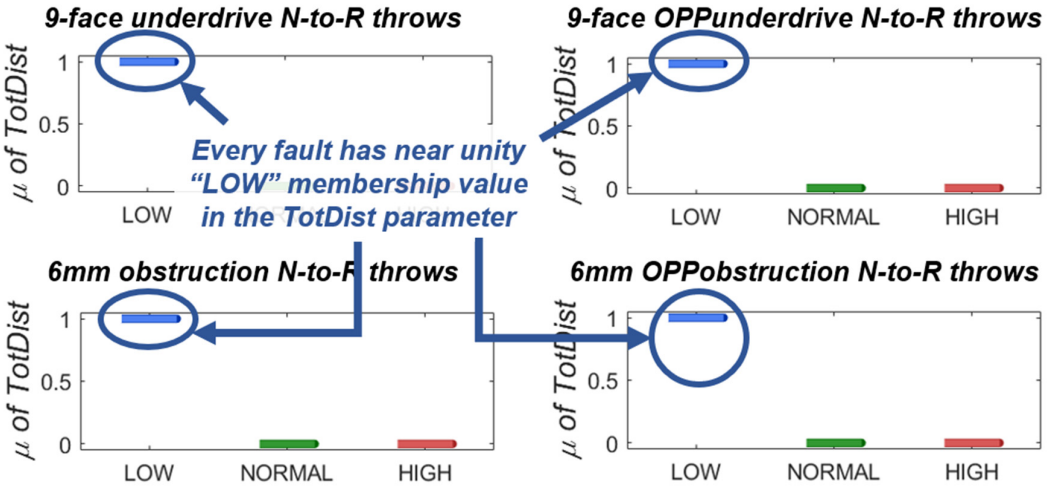


Figure 6-5 Distributions of the three TotDist membership values from all four fault conditions that reduce the switch rail actuation stroke significantly

Second, "9-face overdrive" conditions were found to cause significant increases in both B_{TO} and K_{TO} estimates due to the excessive hard press at the end of the throws.

This also led *TotDist* to be greater than those of the fault-free operations. The other characteristic appeared in the $\omega_m(t_0)$ estimates which always had negligible "High" membership values. The distributions of all membership values of B_{TO}, K_{TO}, $\omega_m(t_0)$, and *TotDist* of the "9-face overdrive" throws are summarised in Figure 6-6, which can facilitate the formulation of the fuzzy rule for severe overdrive fault as follows:

If	B_{TO}	is	High	AND	K_{TO}	is	High
AND	*TotDist*	is	High	AND	$\omega_m(t_0)$	is	NOT High
Then	State	is	severe Overdrive				(6-2)

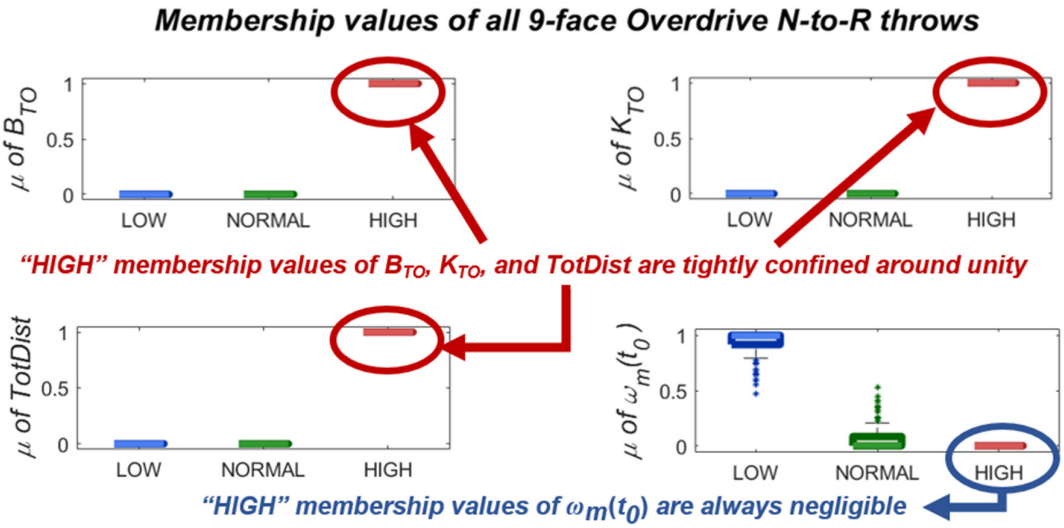


Figure 6-6 Unique effects from the N-to-R 9-face overdrive fault scenarios

The opposite "9-face unlock difficulty" conditions also had a trend of changes in *TotDist* in line with the overdrive scenarios. Besides, this fault substantially increased the J_{PM} estimates, resulting in them always having significant "High" membership values. However, as the distribution of B_{PM} was more dispersed, a suitable conclusion would be them having negligible "Low" membership values instead. The distributions of all three membership values of J_{PM}, B_{PM}, and *TotDist* of the "9-face unlock difficulty" throws are plotted in Figure 6-7, with the associated fuzzy rule as follows:

If J_{PM} is High AND B_{PM} is NOT Low
AND *TotDist* is High
Then State is severe Unlock Difficulty (6-3)

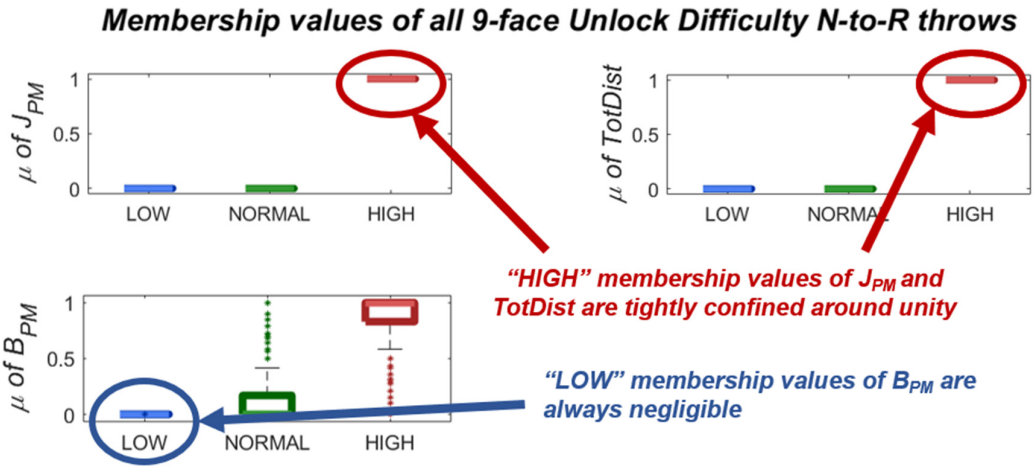


Figure 6-7 Unique effects from the N-to-R 9-face unlock difficulty fault scenarios

Lastly, the abnormal input voltage fault simulations influenced changes in all turnout parameters. For "120 V overvoltage" faults, the SRIVC estimation algorithm perceived that the operations were less restricted. Therefore, most turnout parameter estimates would be lower than those in fault-free conditions, but the $\omega_m(t_0)$ estimates showed the opposite trend instead. In contrast, the algorithm perceived that the turnout operation was more resistive in the "90 V undervoltage" scenarios. So, the effects on all turnout parameters would be the complete opposite of those that appeared in the "120 V overvoltage" faults. Similarly, considering distributions of the three membership values of all affected parameters, the fuzzy rules to detect the occurrence of severe overvoltage and severe undervoltage conditions can be expressed as follows:

If	R_A	is	High	AND	L_A	is	High	
AND	k_M	is	High	AND	J_{PM}	is	High	
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	High	
AND	B_{TO}	is	High	AND	K_{TO}	is	NOT Low	
AND	$\omega_m(t_0)$	is	Low	AND	*TotDist*	is	NOT High	
Then	State	is	severe Undervoltage					(6-4)

If	R_A	is	NOT High	AND	L_A	is	NOT High	
AND	k_M	is	Low	AND	J_{PM}	is	Low	
AND	B_{PM}	is	NOT High	AND	J_{TO}	is	Low	
AND	B_{TO}	is	Low	AND	K_{TO}	is	NOT High	
AND	$\omega_m(t_0)$	is	High	AND	*TotDist*	is	NOT Low	
Then	State	is	severe Overvoltage					(6-5)

In this thesis, the membership value of each fault condition is considered to be equal to the firing strength of the associated rule. Therefore, the “AND” fuzzy operator previously used in all rules will be translated into the minimum function, while the “NOT” fuzzy operator will be translated into the complement of that fuzzy set. For example, the firing strength of the severe Overvoltage rule can be calculated as follows:

$$\begin{aligned} \textit{Firing strength of severe Overvoltage} = \text{MIN}([1-R_{A_}\mu\text{High}], \\ [1-L_{A_}\mu\text{High}],\ k_{M_}\mu\text{Low},\ J_{PM_}\mu\text{Low},\ [1-B_{PM_}\mu\text{High}], \\ J_{TO_}\mu\text{Low},\ B_{TO_}\mu\text{Low},\ [1-K_{TO_}\mu\text{High}],\ \omega_m(t_0)_{_}\mu\text{High}, \\ [1-\textit{TotDist}_{_}\mu\text{Low}]) \end{aligned} \tag{6-6}$$

The membership values of all other severe fault conditions can be computed by the same principle. Based on such a consideration of the unique effects on estimation results caused by all severe fault conditions, the “fault-free” fuzzy rule developed from the complement of severe fault conditions can be formulated as follows:

If (State is NOT Severe_Overdrive) AND

(State is NOT Severe_UnlockDifficulty) AND

(State is NOT Severe_OverVoltage) AND

(State is NOT Severe_UnderVoltage) AND

(State is NOT Severe_Obstruction) AND

(State is NOT Severe_OPP-Obstruction) AND

(State is NOT Severe_Underdrive) AND

(State is NOT Severe_OPP-Underdrive)

Then State is Fault-free. (6-7)

With this fault detection rule, the membership value of the "fault-free" fuzzy set extracted from 206 fault-free and 354 improper sliding surface N-to-R throws can be plotted, as shown in Figure 6-8. The latter is included since it was previously shown in Table 5-5 and Figure 5-34 that *the parameter estimates extracted from the simulated improper sliding surface operations are not different to the fault-free ones.* To further verify the effectiveness of this fault detection rule, all other faulty throws are also evaluated and their "fault-free" membership values are plotted as histograms in the bottom panel of the same figure.

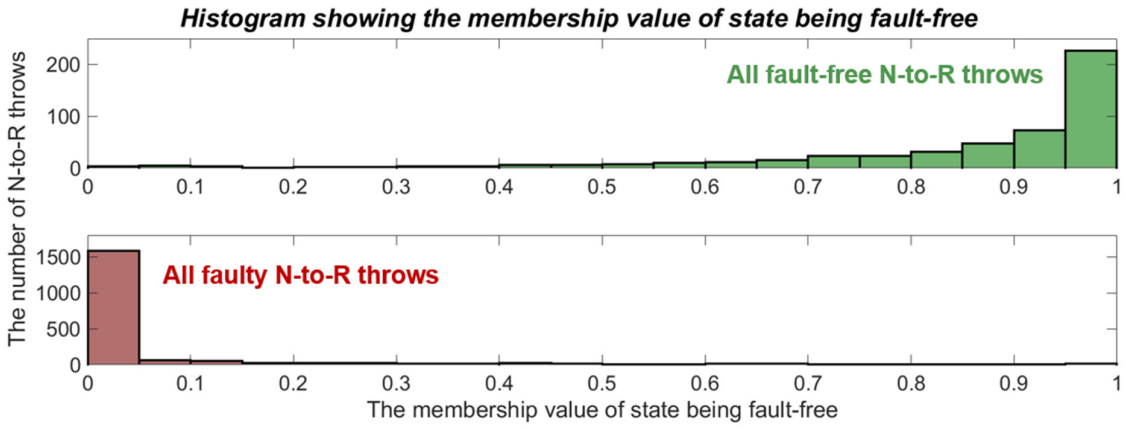


Figure 6-8 Histograms of the "fault-free" membership values of all N-to-R fault-free throws (top) and all N-to-R faulty throws (bottom) calculated from the fault detection rule

The figure illustrates that a substantial 93.08% of the “fault-free” membership values extracted from all N-to-R fault-free operations are greater than 0.5, with a few exceptions that can be considered outliers. Meanwhile, for all faulty throws, their “fault-free” membership values are tightly confined around 0. The ones that possess higher values were found to be less severe scenarios, whereas ones that have absolutely zero “fault-free” membership values were more severe. In conclusion, the procedure of utilising the input fuzzification results to formulate the fault detection rule has been demonstrated in this section. The following section focuses on the fault diagnosis rules required for classification and how the FDD results are interpreted.

6.2.4 Fault diagnosis rules and interpretation of FDD results

The result from the previous rule was only to obtain the “fault-free” membership value of each operation. In this sub-section, the individual membership value of each fault mode will be determined, which can be interpreted as the “likelihood” of that fault occurring. The algorithm is structured to ensure that the sum of membership values from all health states (fault-free + eight detectable fault conditions) still equals unity. As a result, the membership value of each fault mode must be normalised as follows:

$$\text{state}_\mu\text{Fault}(i)_F = \frac{\text{state}_\mu\text{Fault}(i)_I}{\sum_{j=8}\text{state}_\mu\text{Fault}(j)_I} \times \text{state}_\mu\text{Faulty} \qquad (6\text{-}8)$$

where state_μFault(i)$_I$ is the initial membership value of i^{th} fault before normalisation, which equals the firing strength of the associated fault diagnosis fuzzy rule;

state_μFault(i)$_F$ is the final membership value of i^{th} fault after normalisation;

state_μFaulty is the “faulty” membership value that can be calculated from the complement of the “fault-free” membership value, i.e. 1–state_μFaultFree;

Fault(i) represents each simulated fault mode, excluding improper sliding surface faults as the estimation algorithm could not detect significant changes in turnout parameters. So, it only considers is a total of eight fault modes.

The following paragraphs provide three examples to elaborate on how the fault diagnosis rules are formulated. As the continuing theme of this algorithm development, thorough traceability will be maintained by formulating these fuzzy rules based on the relationships between the qualitative changes in turnout parameters and associated fault modes previously provided in Table 5-5. First, the fuzzy rule that can be used to identify the overdrive fault occurrences can be summarised as in Eq. (6-9):

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Overdrive						↑	↑↑	↑↑	↓	↑↑

Fuzzy rule to detect the occurrence of "overdrive" faults

If	R_A	is	ANY	AND	L_A	is	ANY	
AND	k_M	is	ANY	AND	J_{PM}	is	ANY	
AND	B_{PM}	is	ANY	AND	J_{TO}	is	NOT Low	
AND	B_{TO}	is	High	AND	K_{TO}	is	High	
AND	$\omega_m(t_0)$	is	NOT High	AND	*TotDist*	is	High	
Then	State	is	Overdrive					(6-9)

In the table, a shaded grey cell indicates that that associated parameter is not affected by the overdrive fault, hence its contribution to the fuzzy rule is negligible. Meanwhile, since B_{TO}, K_{TO}, and *TotDist* were greatly impacted by this fault and their numerical estimates were higher when the fault got more severe, their fuzzy variables would possess a significant membership value in the "High" fuzzy set. In contrast, the $\omega_m(t_0)$ estimates became lower instead. However, the effects on this parameter were less dominant than those for the previous parameters. As a result, the contribution of $\omega_m(t_0)$ in the fuzzy rule would be that it should not possess a considerable "High" membership value instead. The same principle can also be applied to the J_{TO} estimates, such that their "Low" membership value should be negligible.

Another example that sees the changes in all turnout parameters comes from the undervoltage faults. Its associated fuzzy rule can be summarised as in Eq. (6-10):

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Undervoltage	↑	↑	↑↑	↑↑	↑	↑	↑↑	↑	↓↓	↓

Fuzzy rule to detect the occurrence of "undervoltage" faults

If	R_A	is	NOT Low	AND	L_A	is	NOT Low	
AND	k_M	is	High	AND	J_{PM}	is	High	
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	NOT Low	
AND	B_{TO}	is	High	AND	K_{TO}	is	NOT Low	
AND	$\omega_m(t_0)$	is	Low	AND	*TotDist*	is	NOT High	
Then	State	is	Undervoltage					(6-10)

The last example illustrates the underdrive faults that are almost inseparable from the fault-free operations, except for the effects on the parameters J_{TO} and *TotDist*. The fuzzy rule used to identify their occurrence can be summarised in Eq. (6-11):

Fault modes	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Underdrive						↓				↓↓

Fuzzy rule to detect the occurrence of "underdrive" faults

If	R_A	is	ANY	AND	L_A	is	ANY	
AND	k_M	is	ANY	AND	J_{PM}	is	ANY	
AND	B_{PM}	is	ANY	AND	J_{TO}	is	NOT High	
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY	
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	Low	
Then	State	is	Underdrive					(6-11)

The other five fault diagnosis fuzzy rules are provided in Appendix 3, which are all formulated based on the same principle previously applied to these examples.

Once the unique relationships between the turnout parameters and each fault mode are taken into account, the overlaps in fuzzy rules must be identified and handled to prevent conflicting FDD results. First, it is found that the “undervoltage” rule overlaps with both fuzzy rules of obstruction-related faults. Due to the fact that the conditions in the “undervoltage” rule are more restricted (affecting every parameter), it will have a higher hierarchy. So, the rules of obstruction-related faults must be modified as follows:

Modified rule to detect the occurrence of “obstruction” faults

If	State	is	NOT UnderVoltage	
AND	State	is	Obstruction	
Then	State	is	Obstruction	(6-12)

Modified rule to detect the occurrence of “OPP-obstruction” faults

If	State	is	NOT UnderVoltage	
AND	State	is	OPP-Obstruction	
Then	State	is	OPP-Obstruction	(6-13)

What is more, this overlapping issue appears between the obstruction-related and underdrive-related faults as well. Since the fuzzy rules of the latter are the most relaxed ones, the obstruction-related rules take priority. Therefore, the fuzzy rules for identifying both underdrive-related faults must be modified as follows:

Modified rule to detect the occurrence of “underdrive” faults

If	State	is	NOT UnderVoltage	
AND	State	is	NOT Obstruction	
AND	State	is	NOT OPP-Obstruction	
AND	State	is	Underdrive	
Then	State	is	Underdrive	(6-14)

Modified rule to detect the occurrence "OPP-underdrive" faults

If	State	is	NOT UnderVoltage	
AND	State	is	NOT Obstruction	
AND	State	is	NOT OPP-Obstruction	
AND	State	is	OPP-Underdrive	
Then	State	is	OPP-Underdrive	(6-15)

Once the FIS has been completely formulated, FDD results can be obtained; some examples are displayed in Figure 6-9 to 6-13. In these figures, the top panel shows the membership values of all parameters, aiming to provide raw information that enables skilled personnel to validate the algorithm results based on their experience by using interpretable FDD features. Then, a bar chart indicating the occurrence likelihood of all nine health states is provided in the bottom panel with a textbox that summarises the results into a single sentence based on the maximum likelihood.

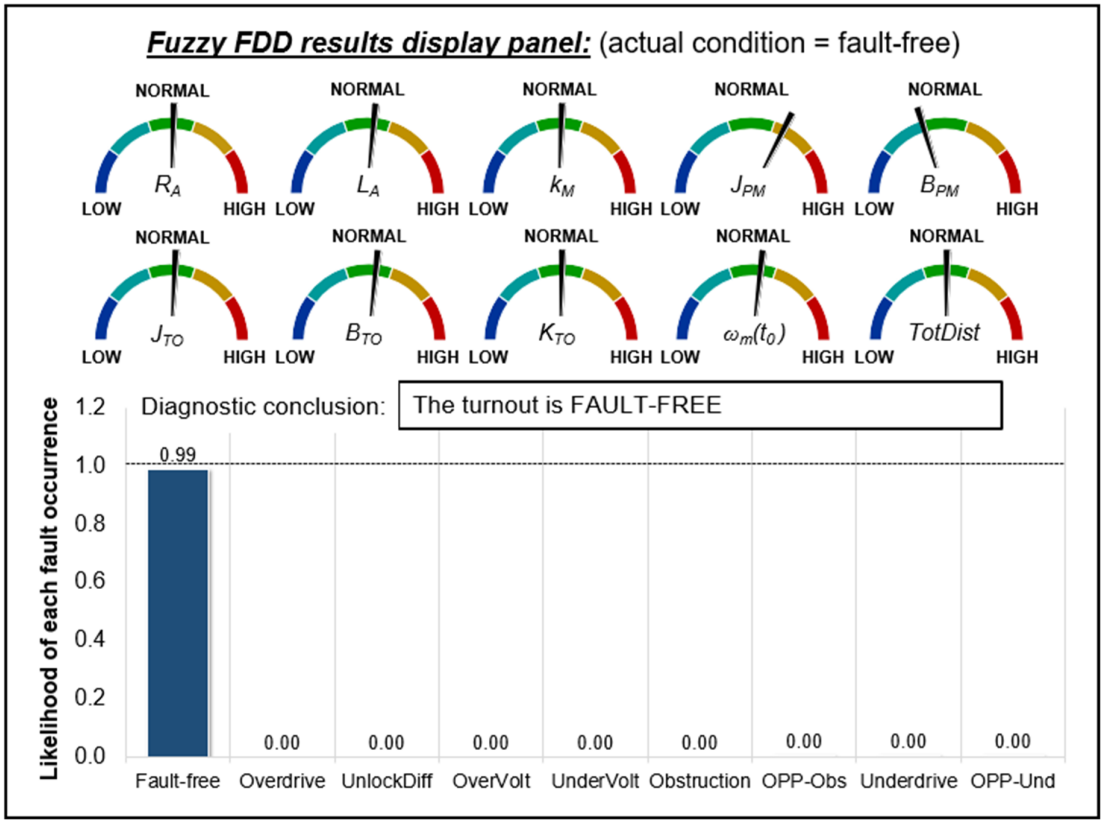


Figure 6-9 Example of FDD result display panel for a fault-free N-to-R throw

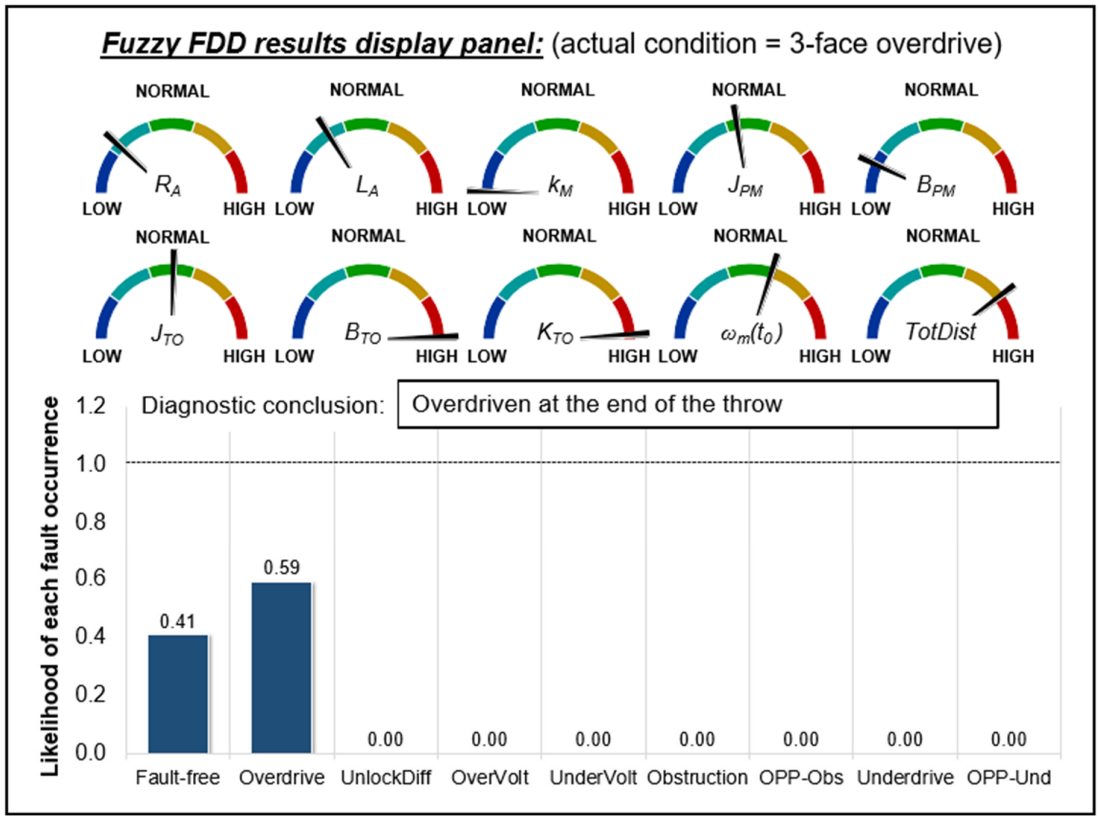


Figure 6-10 Example of FDD result display panel for a 3-face overdrive N-to-R throw

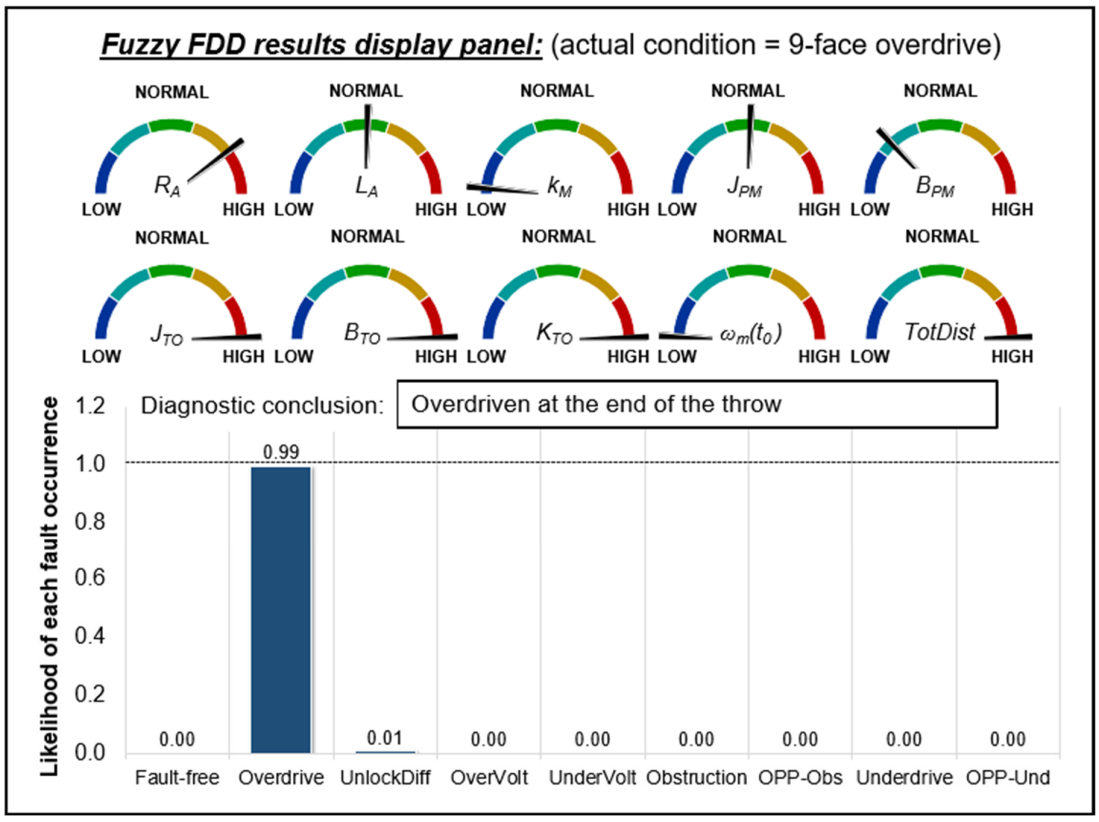


Figure 6-11 Example of FDD result display panel for a 9-face overdrive N-to-R throw

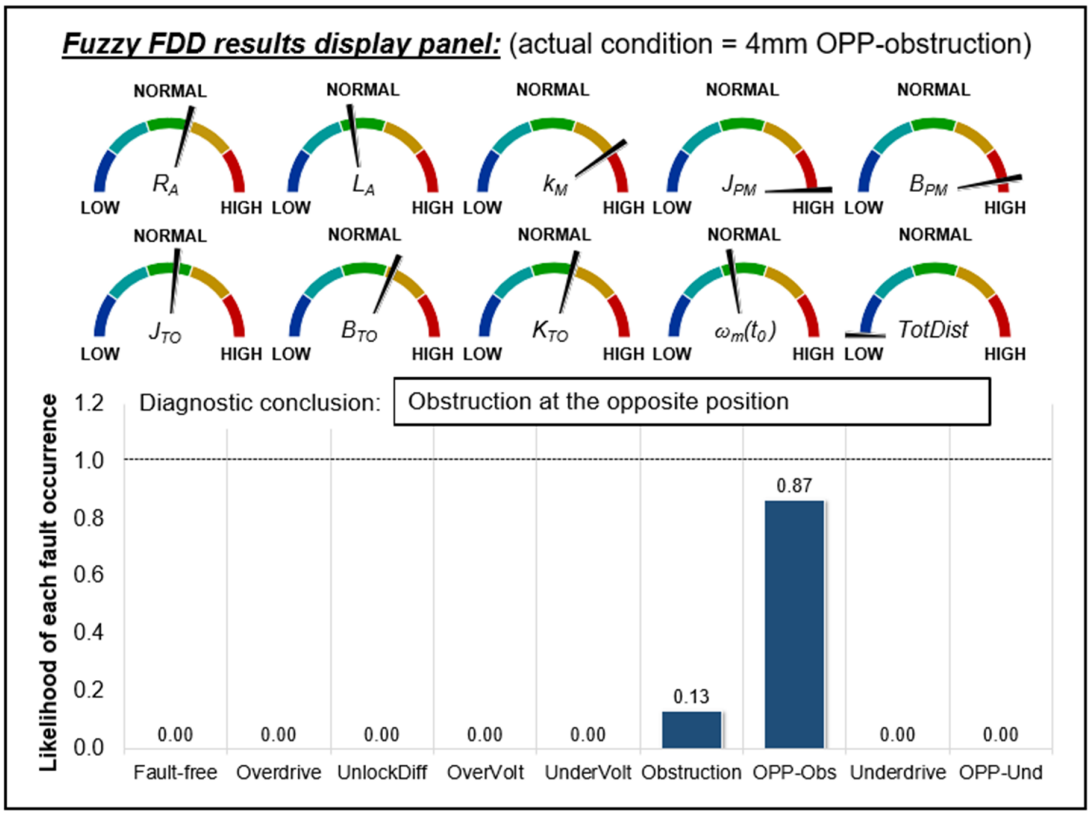


Figure 6-12 Example of FDD result display panel for a 4 mm OPP-obstruction N-to-R throw

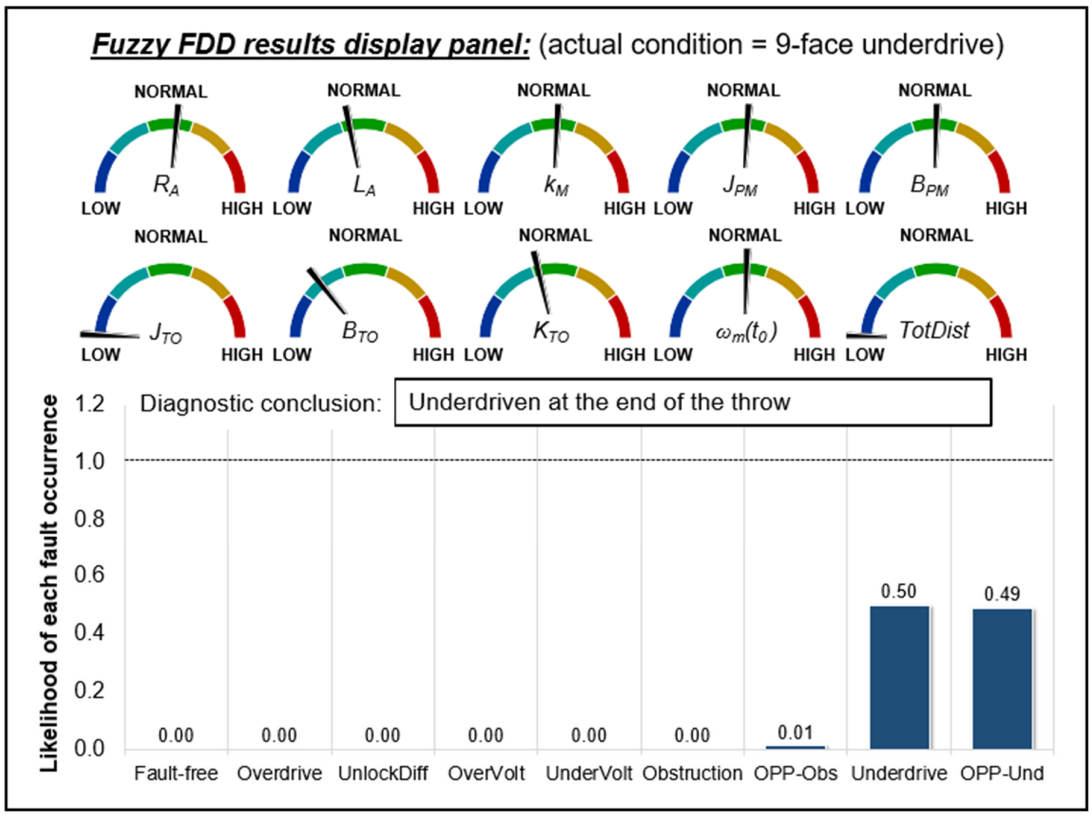


Figure 6-13 Example of FDD result display panel for a 9-face underdrive N-to-R throw

The figures indicate that when the faults get more severe (from Figure 6-10 to 6-11), the likelihood of that fault tends to be higher, and the likelihood of the turnout being fault-free is proportionally reduced. However, for underdrive faults, it is revealed that the algorithm still cannot effectively identify which set of lock nuts was moved from their nominal position. This results in similar occurrence likelihoods of “underdrive” and “OPP-underdrive” faults. Such inseparability in Figure 6-13 is caused by non-different effects resulting from the two fault conditions as perceived by the estimation algorithm.

In this section, the development of the fuzzy logic-based FDD algorithm has been completed. In the following section, the FDD results of the training and testing datasets obtained from the HW55S turnout system will be evaluated to validate the effectiveness of the entire proposed parameter estimation-based FDD methodology.

6.3 Results from the HW55S turnout system

The effectiveness of the proposed FDD methodology is validated in this section. To facilitate the determination of the methodology’s accuracy, a threshold of 0.5 was selected arbitrarily to decide on the nine occurrence likelihoods. Hence, the final FDD result will imply a health state that possesses a likelihood of more than 0.5. If none of the health states meets this requirement, the result will be labelled as “indecisive”.

This threshold is used only to enable a comparison between the accuracy of this approach and that of other methods. The real intention of the proposed methodology is still to provide the occurrence likelihood trends for all health states since they are correlated with their severity, allowing users to observe their development over time.

6.3.1 Results from the training dataset

This section presents the diagnostic results of the training dataset previously used to construct the entire FDD methodology. The collection of 2,495 N-to-R throws

was experimented on ten separate days between August and October 2022. Owing to the fact that the proposed FDD methodology is not effective in isolating the “underdrive” and “OPP-underdrive” conditions (see Figure 6-13), their occurrence likelihoods are added into a single value, only to signify that an underdrive fault has occurred.

The confusion matrix describing the FDD results between the actual condition and the predicted outcome of all N-to-R training operations is illustrated in Figure 6-14. The cells with a 0% entry are all faded to provide better visualisation, whereas the green cells indicate the FDD results that correctly classify the health state.

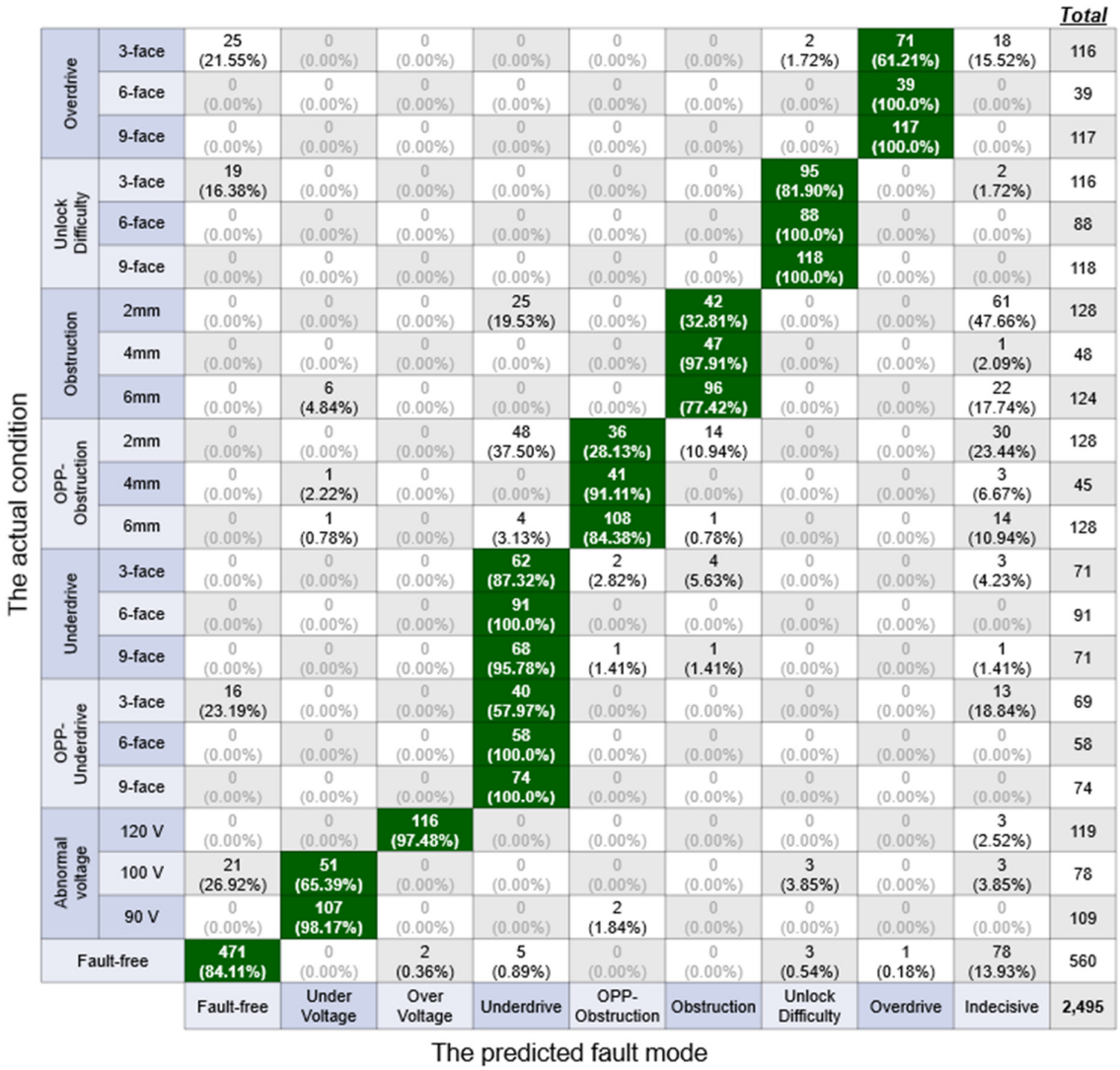


Figure 6-14 Confusion matrix showing the FDD results of the N-to-R HW55S turnout training dataset produced by the proposed methodology with the likelihood threshold at 0.5

The figure shows that the FDD methodology achieves 81.60% overall accuracy (i.e. correctly identifies 2,036 operations out of the total 2,495). Particularly for each fault condition, when faults become more severe, the methodology tends to classify them more accurately. In contrast, for less severe fault conditions, the methodology produces some “indecisive” results and misclassifications. The misclassified results are mostly found between obstruction and underdrive-related faults and there are also a few between obstruction-related and undervoltage faults. This is expected since both pairs were previously identified to produce effects that overlapped with each other.

Nonetheless, the diagnosis of an individual operation alone is inadequate for effective monitoring, since another important piece of information also lies in the trend of the turnout deterioration and fault development. To provide better-visualised trends of the likelihood of a fault occurring, which strongly correlates with fault development, a moving average filter with a window size of 10 samples is applied to the fuzzy diagnostic results, similar to the approach previously used by Silmon (2009).

Examples of the occurrence likelihood trends of all nine health states extracted from the N-to-R training dataset are illustrated in Figure 6-15. At the top of each graph, there is information on the actual conditions being simulated on the HW55S turnout during that period, which can also be used to assess the algorithm’s performance. The abbreviations of each fault condition used in Figure 6-15 are as follows:

FF	: Fault-free	OBS	: Obstruction
OD	: Overdrive	OPOb	: OPP-Obstruction
UnDif	: Unlock Difficulty	UD	: Underdrive
OV	: Overvoltage	UV	: Undervoltage

Note that the confusion matrix of the R-to-N HW55S training operations and their associated occurrence likelihood trends is also provided in Appendix 4.

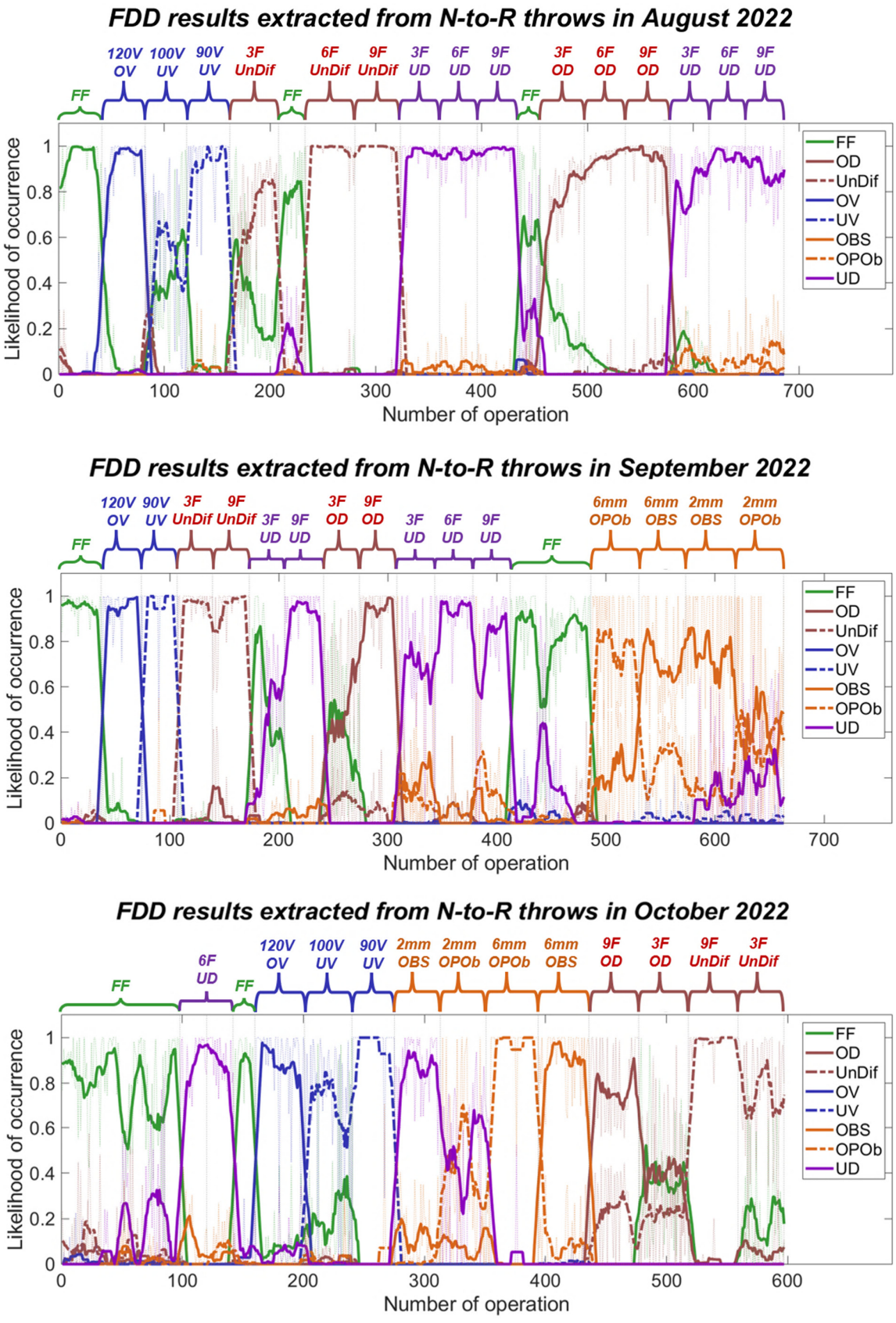


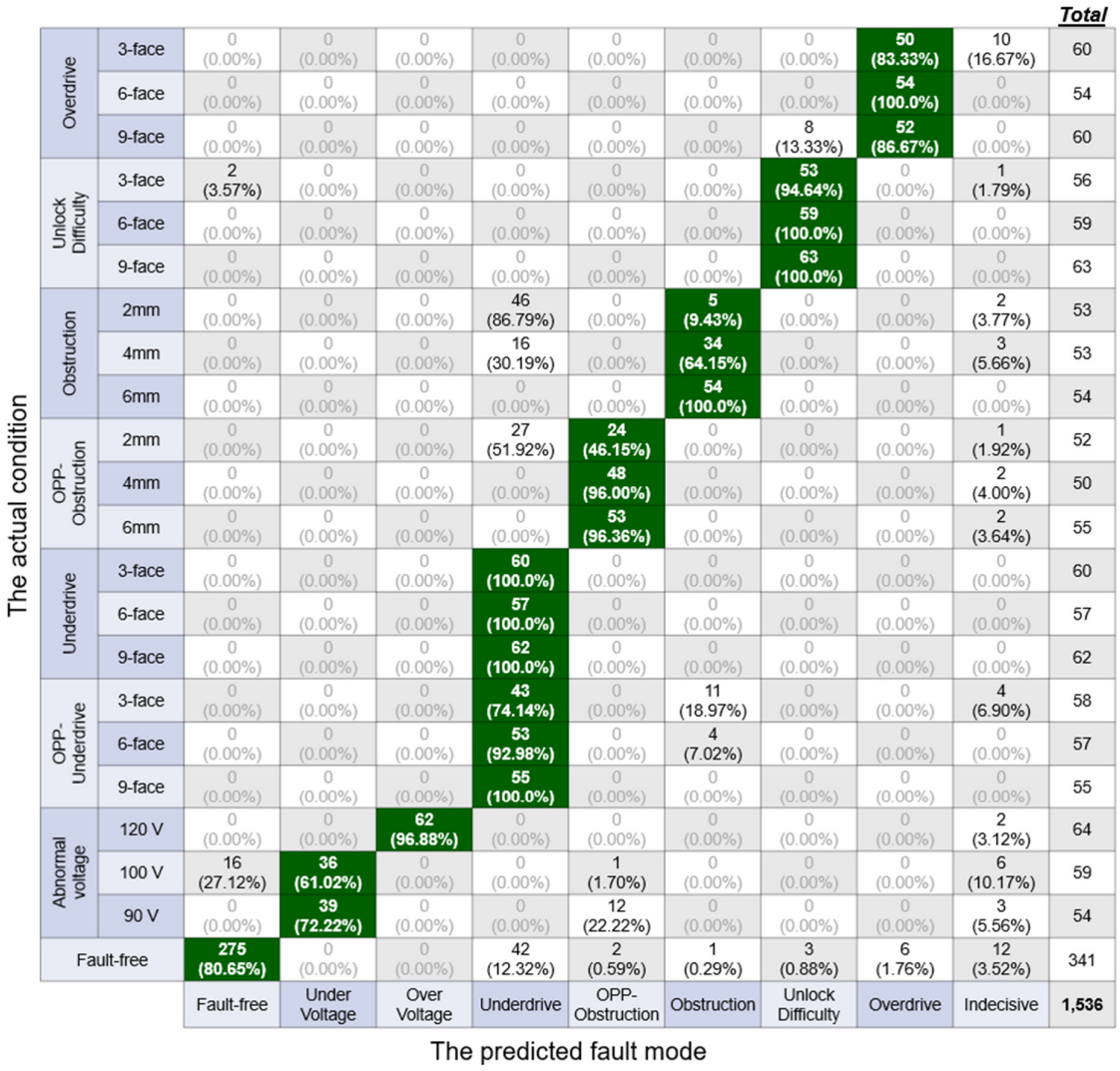
Figure 6-15 Examples of occurrence likelihood trends extracted from the training dataset

Figure 6-15 indicates that the fault occurrence likelihood trends correspond well with the actual conditions. Taking the experiments in August as an example, the first 41 throws were fault-free, so the “fault-free” likelihood is consistently more than 0.8. Then, the 120 V overvoltage conditions were simulated on the following 41 operations. The likelihood of the turnout being “fault-free” is now significantly reduced, while that of it being “overvoltage” increases to almost unity. Finally, the subsequent 39 throws were then simulated with 100 V undervoltage conditions instead. Seeing that it is not a severe undervoltage fault, the likelihood of the turnout system being “fault-free” escalates again together with that of it being “undervoltage”, whereas the likelihood of the turnout system being “overvoltage” drops to zero.

The FDD results obtained in September and October 2022 also follow a similar pattern. From outputting less deterministic FDD results when faults are in their early stages to being more decisive when faults become critical, these likelihood trends allow the maintainers to be notified of ongoing fault developments. To completely verify the effectiveness of the proposed FDD methodology, it was applied to another testing dataset obtained from the same HW55S turnout system with the same experimental procedures, but conducted during February and April 2023.

6.3.2 Results from the testing dataset

Once the proposed FDD methodology had been developed and verified with the training dataset, measurements from another 3,081 throws were used for validation. The same FDD methodology was also applied to this testing dataset, i.e. extracting the turnout parameter estimates from the raw measurements, transforming them into fuzzy variables, and employing the same FIS to produce the fault occurrence likelihoods. First, the confusion matrix providing the diagnostic results of the 1,536 testing N-to-R throws with the occurrence likelihood threshold at 0.5 is given in Figure 6-16.



The actual condition		Fault-free	Under Voltage	Over Voltage	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	Total
Overdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**50 (83.33%)**	10 (16.67%)	60
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**54 (100.0%)**	0 (0.00%)	54
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	8 (13.33%)	**52 (86.67%)**	0 (0.00%)	60
Unlock Difficulty	3-face	2 (3.57%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**53 (94.64%)**	0 (0.00%)	1 (1.79%)	56
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**59 (100.0%)**	0 (0.00%)	0 (0.00%)	59
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**63 (100.0%)**	0 (0.00%)	0 (0.00%)	63
Obstruction	2mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	46 (86.79%)	0 (0.00%)	**5 (9.43%)**	0 (0.00%)	0 (0.00%)	2 (3.77%)	53
	4mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	16 (30.19%)	0 (0.00%)	**34 (64.15%)**	0 (0.00%)	0 (0.00%)	3 (5.66%)	53
	6mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**54 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	54
OPP-Obstruction	2mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	27 (51.92%)	**24 (46.15%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (1.92%)	52
	4mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**48 (96.00%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (4.00%)	50
	6mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**53 (96.36%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (3.64%)	55
Underdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**60 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	60
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**57 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	57
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**62 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	62
OPP-Underdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**43 (74.14%)**	0 (0.00%)	11 (18.97%)	0 (0.00%)	0 (0.00%)	4 (6.90%)	58
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**53 (92.98%)**	0 (0.00%)	4 (7.02%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	57
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**55 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	55
Abnormal voltage	120 V	0 (0.00%)	0 (0.00%)	**62 (96.88%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (3.12%)	64
	100 V	16 (27.12%)	**36 (61.02%)**	0 (0.00%)	0 (0.00%)	1 (1.70%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	6 (10.17%)	59
	90 V	0 (0.00%)	**39 (72.22%)**	0 (0.00%)	0 (0.00%)	12 (22.22%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (5.56%)	54
Fault-free		**275 (80.65%)**	0 (0.00%)	0 (0.00%)	42 (12.32%)	2 (0.59%)	1 (0.29%)	3 (0.88%)	6 (1.76%)	12 (3.52%)	341
		Fault-free	Under Voltage	Over Voltage	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	**1,536**

The predicted fault mode

Figure 6-16 Confusion matrix showing the FDD results of the N-to-R HW55S turnout testing dataset produced by the proposed methodology with the likelihood threshold at 0.5

In this case, the proposed FDD methodology can correctly classify 1,291 testing N-to-R operations, equivalent to 84.05% overall accuracy. Improvement in accuracy for the testing dataset can be explained by the fact that fewer operations were evaluated and they were obtained from only three days. Therefore, the measurements collected from each fault condition tended to be more consistent compared to those of the training dataset. The fault occurrence likelihood trends of the same testing N-to-R operations are then plotted in Figure 6-17, whereas the diagnostic results of the testing R-to-N operations are provided in Appendix 4 as well.

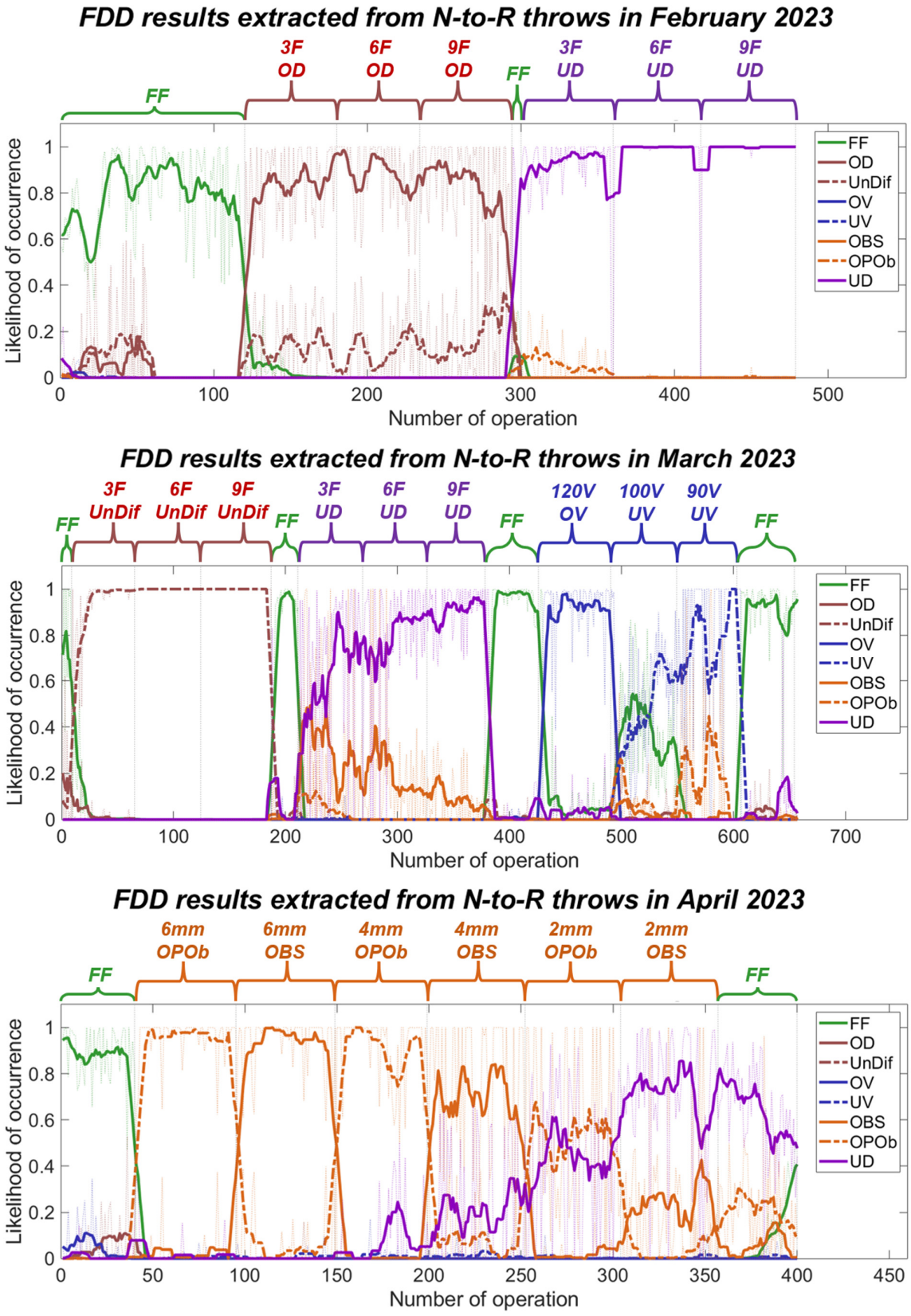


Figure 6-17 Fault occurrence likelihood trends of the entire N-to-R HW55S testing dataset

Compared to the verification results from the training dataset, there are similar correlations between the occurrence likelihood trends and the actual fault simulations for the testing dataset as well. The only exception appears in the last “FF” period in April 2023 data, where the methodology implied these throws as being underdriven. This is because the HW55S system has previously been simulated with many severe obstruction faults. As a result, there was still residual built-up stress inside the point machine, causing it not to be able to close the gap entirely, even when the obstruction had been removed. The recovery is shown to start from the 380th throw.

Due to the accuracy of more than 80%, the use of interpretable parameters to produce FDD results, and the ability to classify multiple fault modes with information on their severity progression, if this HW55S turnout system were now in service, this methodology could immediately be used to support its maintenance planning.

Finally, for any validated FDD methodology to be practically applicable, it should be able to be re-trained to effectively diagnose other turnout systems as well. Thus, in the following section, the proposed methodology will be applied to an HW2000 turnout system to validate its transferability. This will illustrate how the inherent interpretability and traceability of the methodology can facilitate the re-training process.

6.4 Transferability validation: HW2000 turnout system

The effectiveness of the proposed methodology has already been validated with the HW55S turnout system. However, when the methodology is to be applied to other turnouts, it must be re-trained based on the new turnout system characteristics first. If a data-driven methodology were developed instead, a great amount of new training data would be required for this re-training process. Additionally, the optimal features and their associated thresholds for evaluation must be re-identified as well.

In reality, collecting such a comprehensive training dataset from in-field turnout systems that must include scenarios from all fault conditions at multiple severity levels is not practical. Combined with complex diagnostic reasoning commonly used in data-driven methodologies, these contribute to the general barrier that prevents these kinds of FDD approaches from being practically applicable to railway turnout systems.

As a result, this section aims to demonstrate how the inherent interpretability and traceability of the proposed parameter estimation-based methodology can simplify the re-training process that is normally sophisticated in data-driven approaches.

6.4.1 Experimental setup of the HW2000 turnout system

A testing HW2000 turnout system is located outdoors at BK3 Training Limited, Wolverhampton, UK. The point machine is connected to a short section of switch rails as illustrated in Figure 6-18. The figure also shows the locations where the encoder and the linear potentiometer are fitted to measure the driving shaft speed and the total switch rail displacement. Also, the data acquisition system used for this data collection campaign is identical to the one used for the HW55S turnout system.

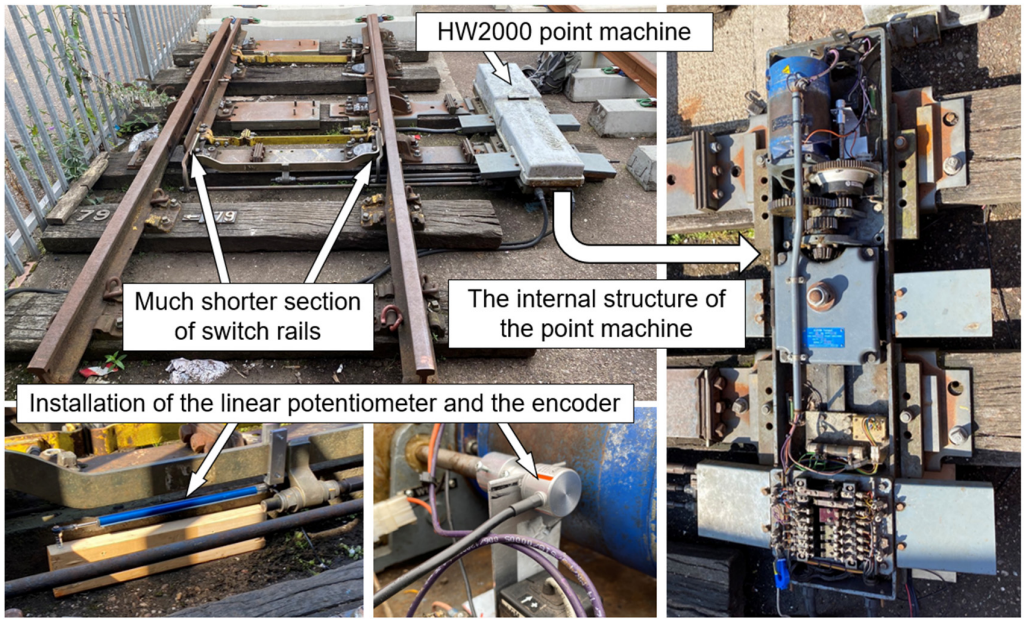


Figure 6-18 Overview of the HW2000 turnout system at BK3 Training Limited

Compared to the HW55S turnout system, the HW2000 uses an electromagnetic clutch, which disengages the motor from the driving mechanism at the end of the throw. This will cause the motor shaft to still spin after the disengagement, while the driving shaft will be brought to a stop suddenly. Also, due to a much shorter section of switch rails being connected to the HW2000 point machine, the entire switch rails will be actuated linearly with little to no bend. Lastly, because the DC power supply is always fixed at 110 V, abnormal input voltage fault simulations cannot be conducted. Apart from these aspects, the working principles of the HW55S and the HW2000 turnout systems are comparable. The next sub-section presents the diagnostic results when the proposed methodology is immediately applied to diagnose the HW2000 system.

6.4.2 Initial FDD results before modification

To assess the transferability of the proposed FDD methodology, when it is first immediately applied to diagnose the HW2000 turnout. Examples of the parameter estimation results from fault-free, simulated “6-face unlock difficulty”, and simulated “4 mm obstruction” N-to-R operations are demonstrated in Figure 6-19 to 6-21.

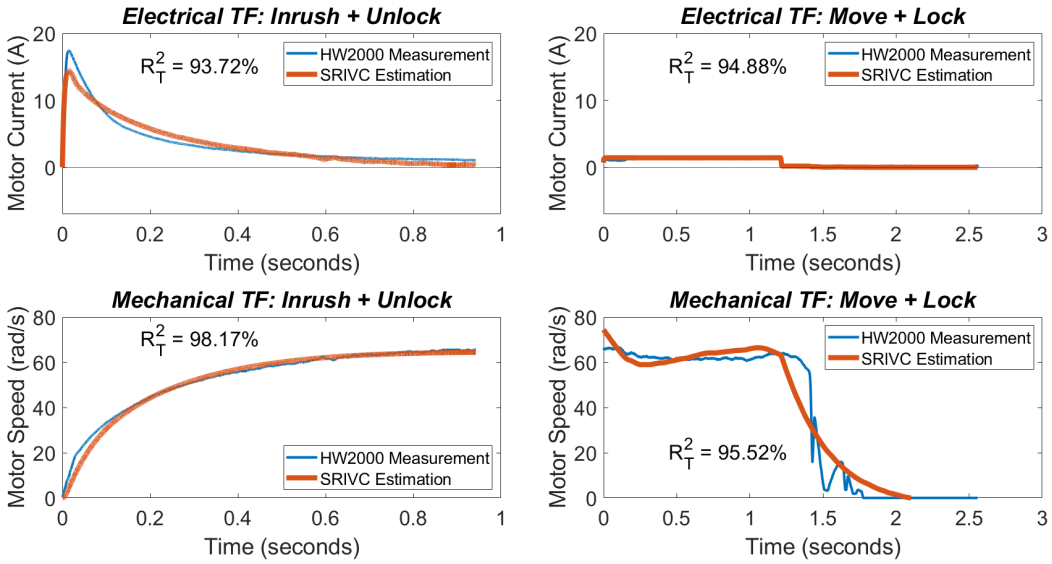


Figure 6-19 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R fault-free throw

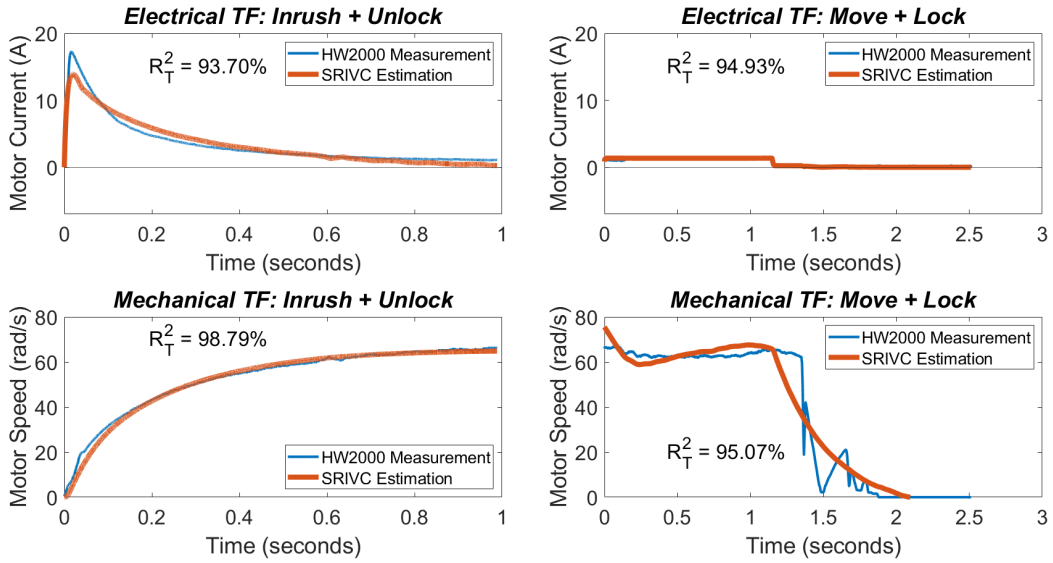


Figure 6-20 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R 6-face unlock difficulty throw

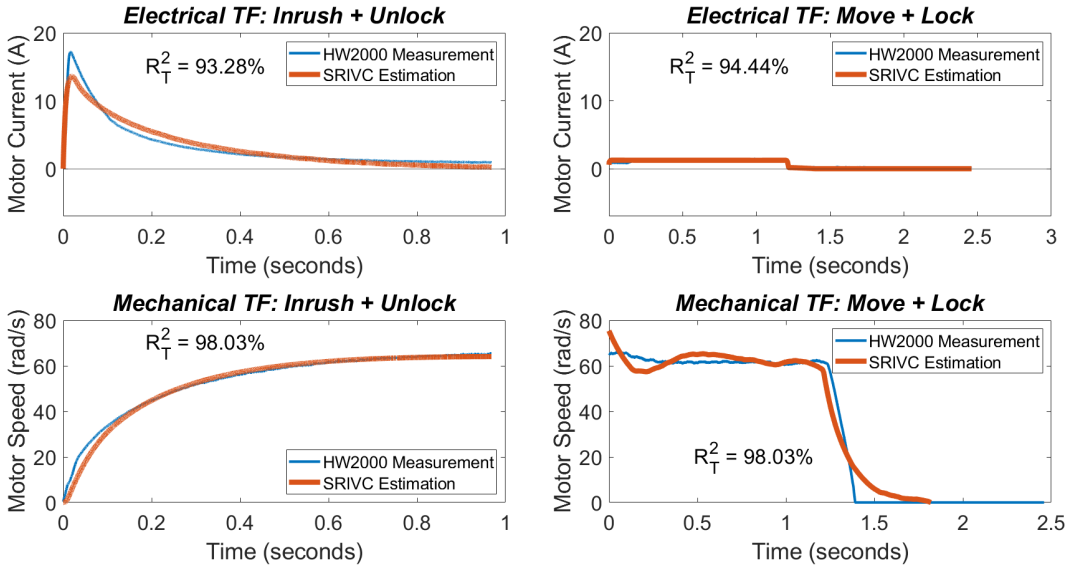


Figure 6-21 Data-fitting performance of the estimation algorithm applied on an example HW2000 N-to-R 4 mm obstruction throw

These figures indicate that the parameter estimation aspect of the methodology is still efficient in obtaining the turnout parameter estimates for the HW2000 system. The following step is to diagnose it by implementing the same fuzzy rule-based FDD algorithm to its physically interpretable parameters. The FDD result display panels of these three N-to-R throws are illustrated in Figure 6-22 to 6-24.

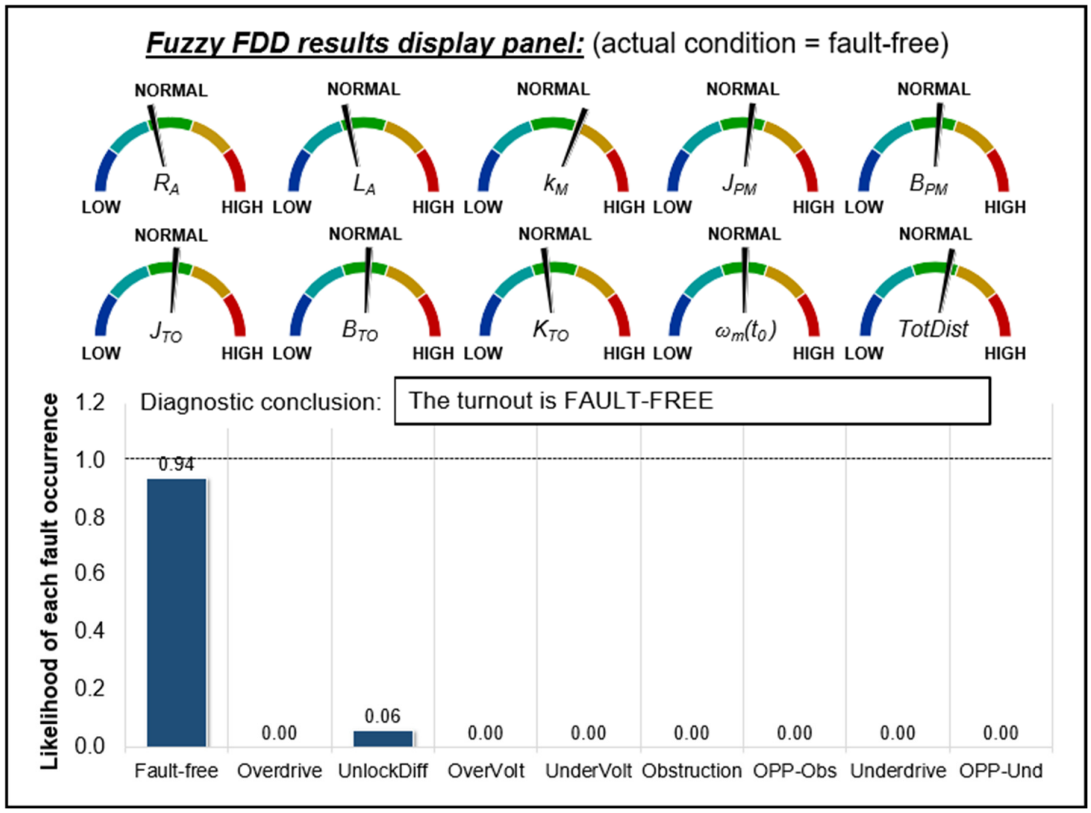


Figure 6-22 FDD display panel of fault-free HW2000 N-to-R throw

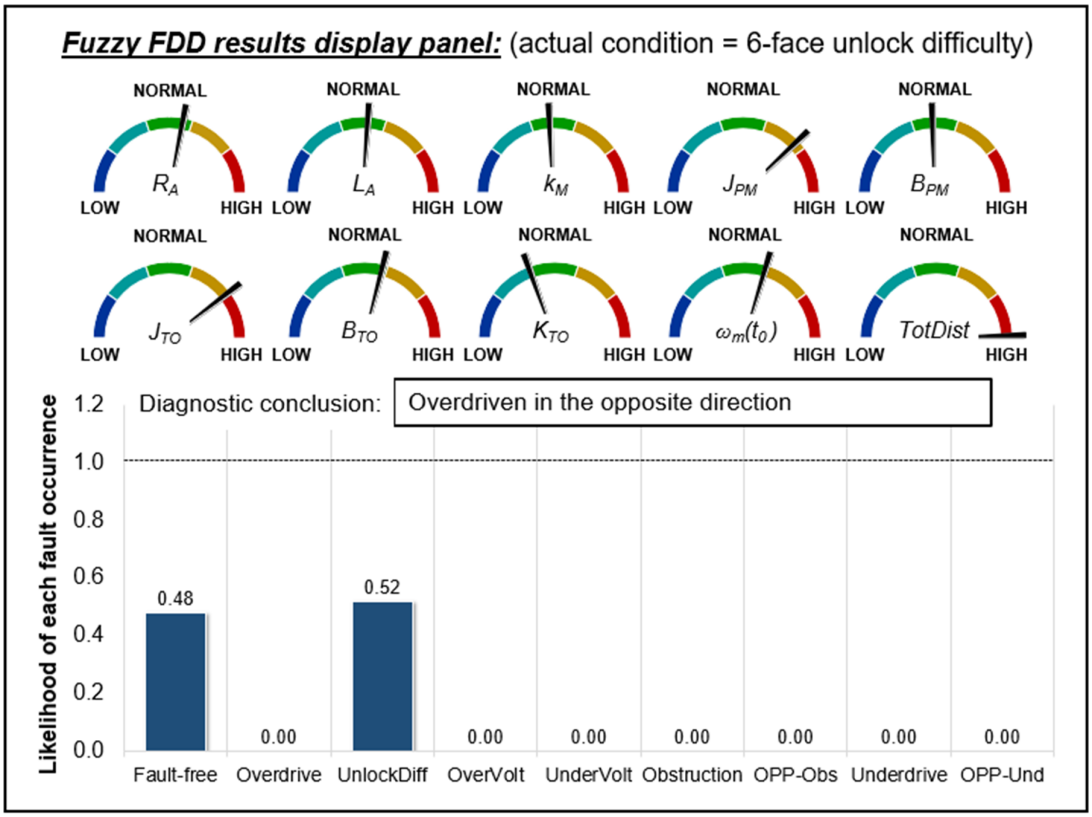


Figure 6-23 FDD display panel of 6-face unlock difficulty HW2000 N-to-R throw

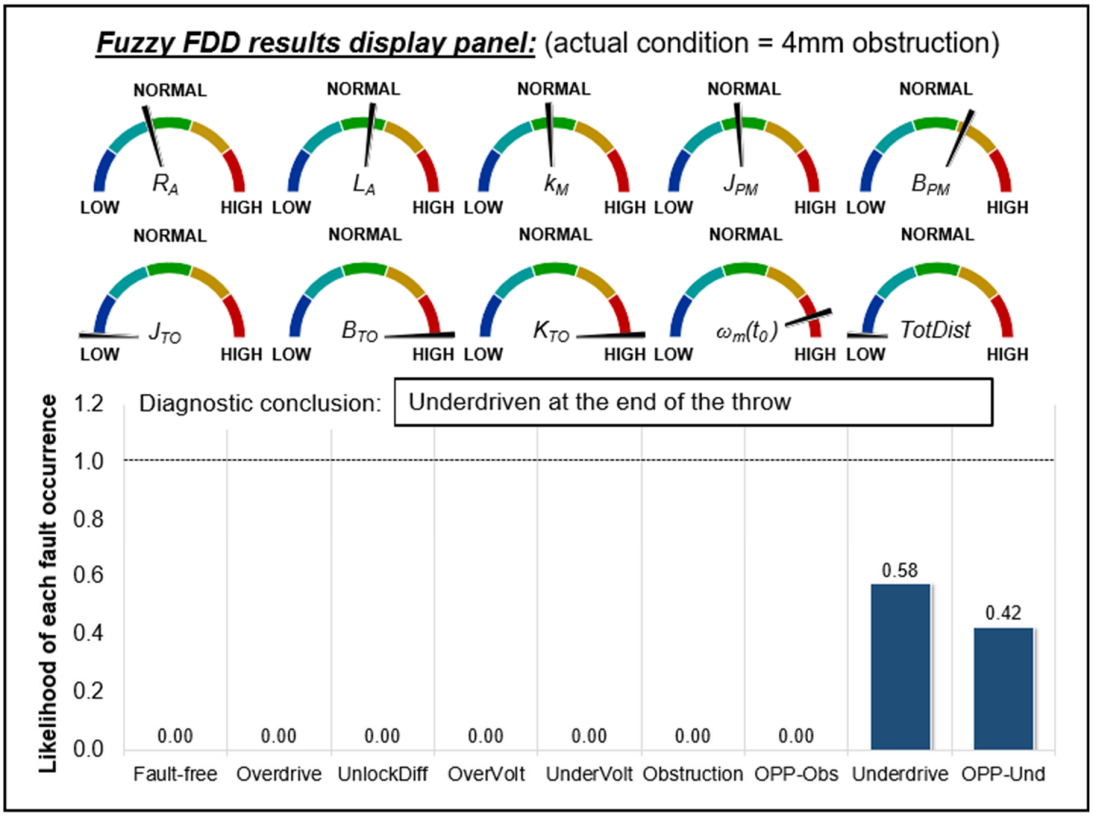


Figure 6-24 FDD display panel of 4 mm obstruction HW2000 N-to-R throw

The FDD result panels in Figure 6-23 and 6-24 demonstrate the ineffectiveness of the proposed methodology when it is immediately applied to the HW2000 system without appropriate modifications. The former reveals that it cannot produce a decisive FDD result even if the fault is severe (i.e. 6-face unlock difficulty condition), whereas the latter displays a misclassification (diagnosed as underdrive instead of obstruction).

This poor performance indicates that the fuzzy rules previously derived from the statistical characteristics of the HW55S turnout estimates are no longer effective. To emphasise the downfalls of the methodology's immediate use, the confusion matrix of all N-to-R HW2000 testing operations with the threshold of occurrence likelihood at 0.5 is shown in Figure 6-25, which specifies that it only achieves 61.41% overall accuracy. Meanwhile, the associated fault occurring likelihood trends extracted from the same N-to-R HW2000 operations are also provided in Figure 6-26.

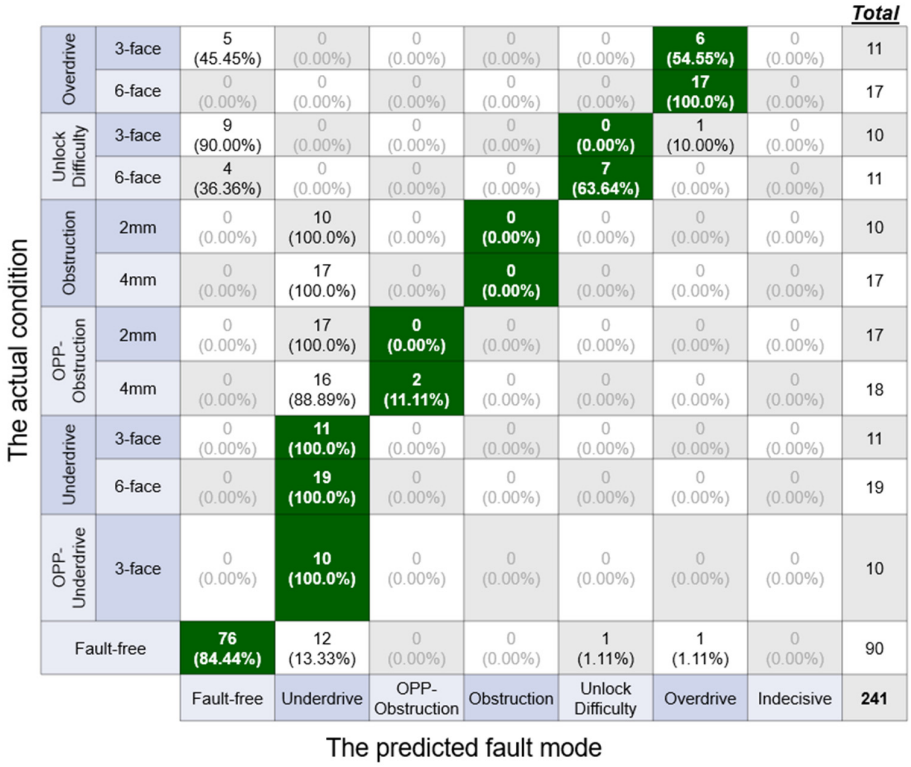


Figure 6-25 Confusion matrix showing the FDD results of the N-to-R HW2000 turnout testing dataset produced by the proposed methodology with the likelihood threshold at 0.5

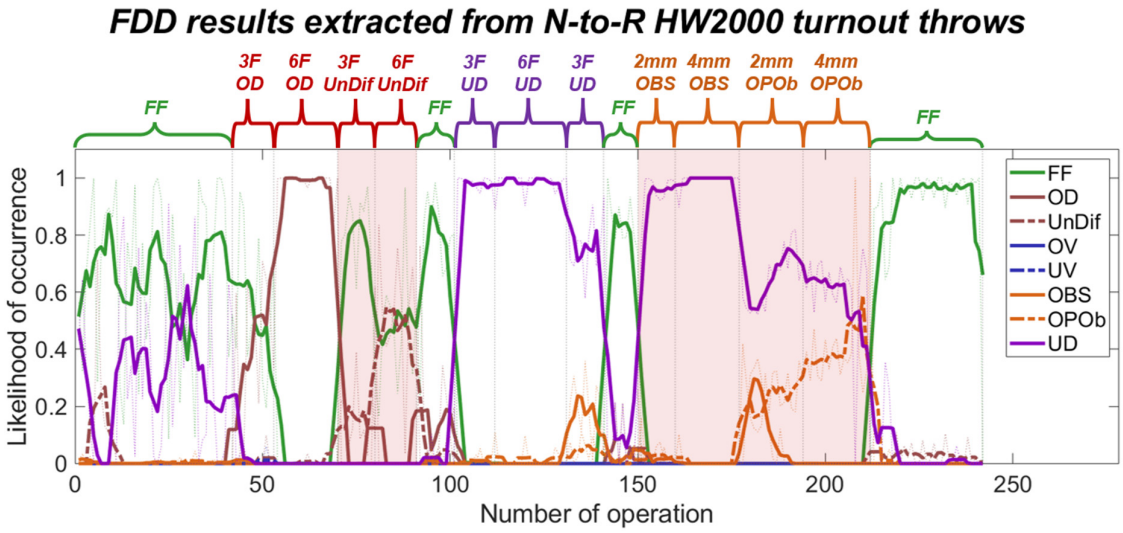


Figure 6-26 Fault occurrence likelihood trends of the entire N-to-R HW2000 testing dataset

These illustrations highlight that to enable the effective FDD capabilities of the proposed methodology when it is used to diagnose the HW2000 turnout system, the fuzzy logic-based FDD classification aspect of the methodology must be modified first.

Therefore, the following section elaborates on the modification process that is facilitated by the inherent interpretability and traceability of the proposed methodology.

6.4.3 Modification of the fuzzy logic-based FDD algorithm

The results presented in the previous section utilized the existing ten turnout parameters, as they retain the same physical interpretations, making them still useful for diagnostics. Also, before the initial application of the proposed FDD algorithm, the HW2000 turnout system was operated in fault-free conditions multiple times. This was necessary to establish baseline values and membership functions for each parameter based on the HW2000's fault-free characteristics, which is equivalent to setting new optimal thresholds for subsequent feature evaluation.

Although all membership functions for each turnout parameter were adjusted during the re-training process, the results from directly applying the proposed algorithm revealed that the fuzzy rules previously used to classify unlock difficulty, obstruction, and OPP-obstruction faults for the HW55S turnout system are no longer suitable for the HW2000, as indicated by the red areas in Figure 6-26. Therefore, these specific fuzzy rules also require modification as part of the re-training process.

Since the fuzzy rules were initially formulated from the statistical characteristics of the HW55S parameter estimates, the same intuitive and traceable method allows these rules to be adapted to the distributions of HW2000 parameter estimates as well. Therefore, a simpler cross-training process for modifying the underperforming fuzzy rules can be illustrated in Figure 6-27 – shown in red. The figure also demonstrates — in orange — that if the system under test is of the same type (e.g. the testing system is another HW55S turnout), only its fault-free characteristics are needed to adjust the baseline values and membership functions for all turnout parameters instead.

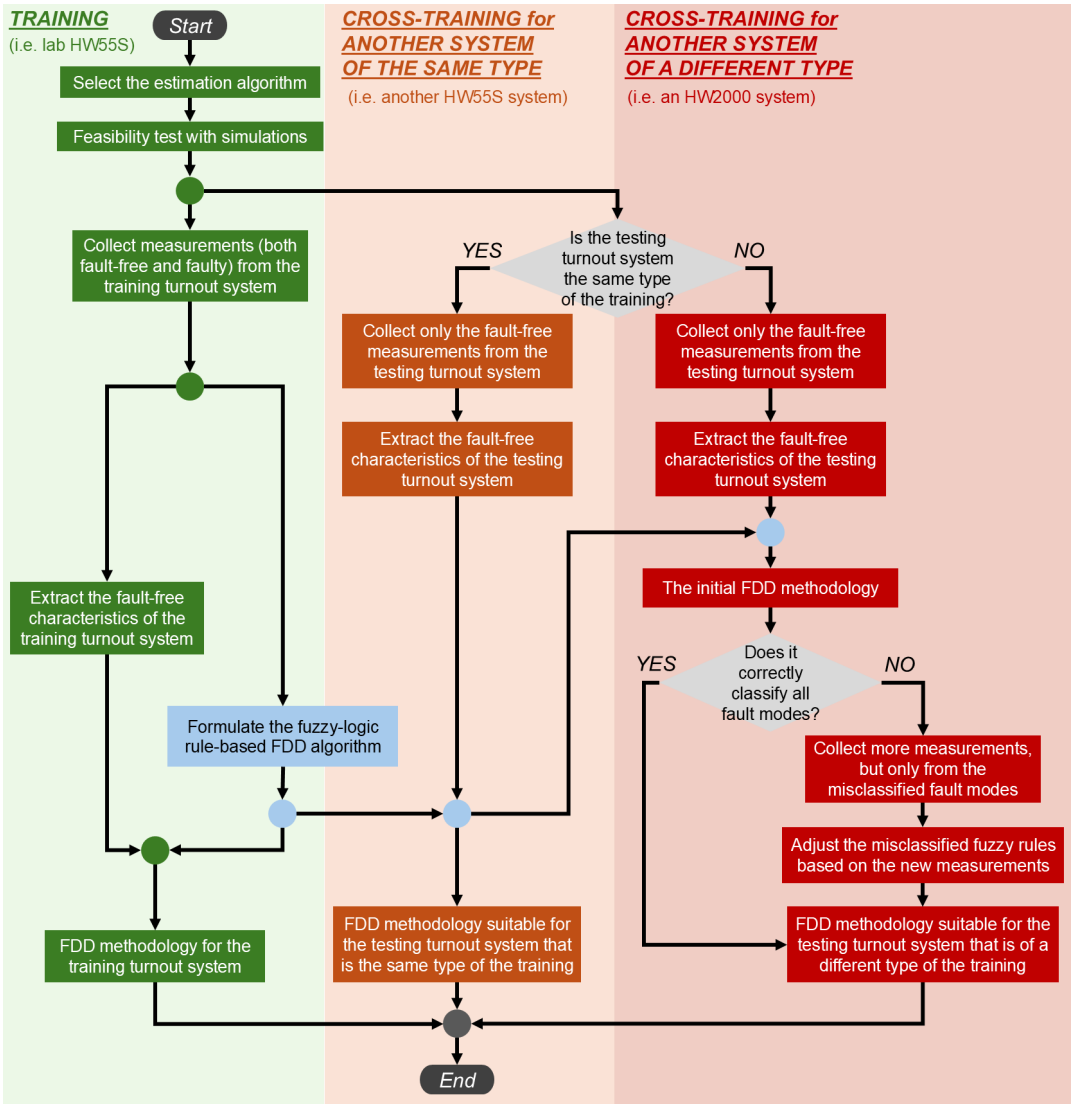


Figure 6-27 The system overview of the proposed FDD methodology, including the cross-training process required for the methodology to be effective for another turnout system

Initially, for unlock difficulty faults, it is found that the membership values of J_{PM}, B_{PM}, and three other parameters obtained from the HW2000 turnout are distributed differently compared to those from the HW55S system, as shown in Figure 6-28. Due to such differences, the associated fuzzy rule must be modified as follows:

Turnout type	***Parameters affected by unlock difficulty conditions***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
HW55S				↑↑	↑					↑↑
HW2000				↑	↓	↑	↑	↓		↑↑

Fuzzy rule to detect "unlock difficulty" faults in HW2000 turnouts

If	R_A	is	ANY	AND	L_A	is	ANY	
AND	k_M	is	ANY	AND	J_{PM}	is	NOT Low	
AND	B_{PM}	is	NOT High	AND	J_{TO}	is	NOT Low	
AND	B_{TO}	is	NOT Low	AND	K_{TO}	is	NOT High	
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	High	
Then	State	is	UnlockDifficulty (for the HW2000 turnout)					(6-16)

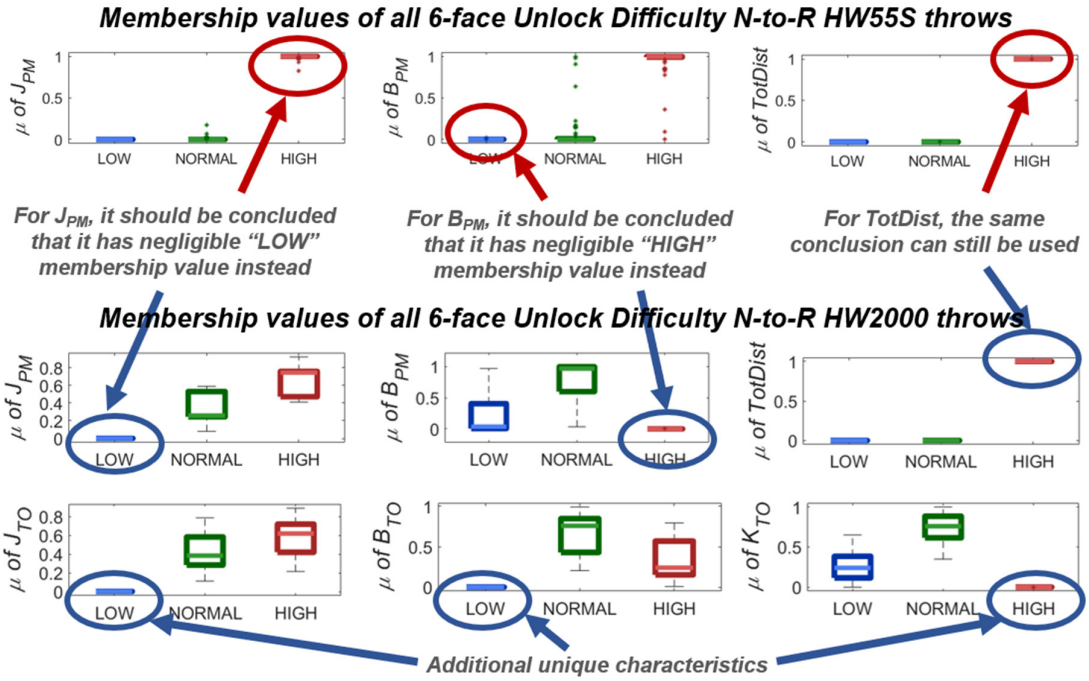


Figure 6-28 Differences between the effects of the 6-face unlock difficulty N-to-R operations on the parameter estimates of HW55S and HW2000 turnout systems

For the obstruction conditions, it is found that both B_{TO} and K_{TO} estimates still hold significant "High" membership values. However, the effects of obstruction faults on the $\omega_m(t_0)$ and J_{TO} estimates extracted from the HW2000 measurements are now different compared to those from the HW55S turnout system due to the differences in the switch rail's structure. These variations are displayed in Figure 6-29. Therefore, similar to how the unlock difficulty fuzzy rule was modified, the rule used to identify the obstruction fault occurrences for the HW2000 turnout is modified as follows:

Turnout type	Parameters affected by obstruction conditions									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
HW55S						↑	↑↑	↑↑	↓	↓↓
HW2000						↓↓	↑↑	↑↑	↑	↓↓

Fuzzy rule to detect "obstruction" faults in HW2000 turnouts

If	R_A	is	ANY	AND	L_A	is	ANY	
AND	k_M	is	ANY	AND	J_{PM}	is	ANY	
AND	B_{PM}	is	ANY	AND	J_{TO}	is	Low	
AND	B_{TO}	is	High	AND	K_{TO}	is	High	
AND	$\omega_m(t_0)$	is	NOT Low	AND	*TotDist*	is	Low	
Then	State	is	Obstruction (for the HW2000 turnout)					(6-17)

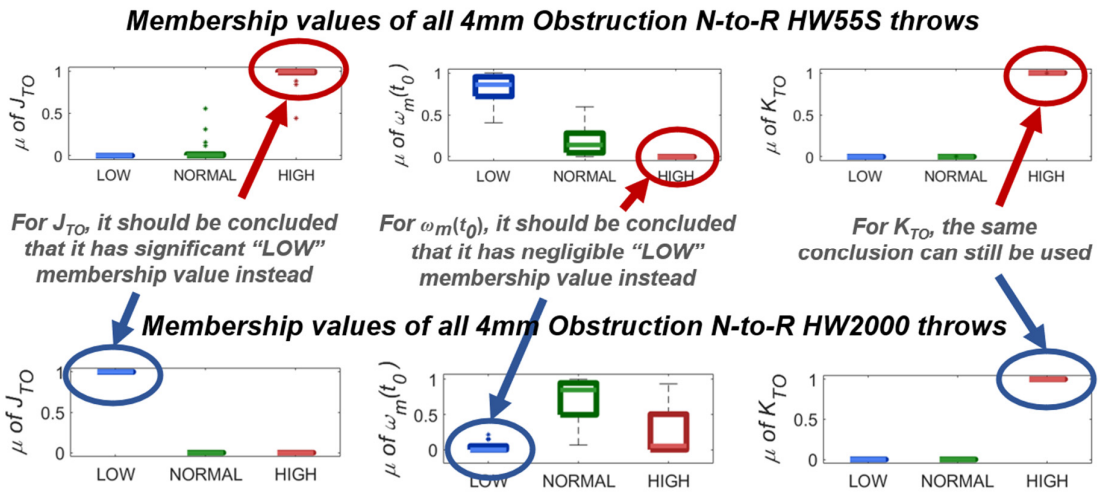


Figure 6-29 Differences between the effects of the 4 mm obstruction N-to-R operations on the parameter estimates of HW55S and HW2000 turnout systems

Finally, for OPP-obstruction faults, with the same modification procedure, the modified fuzzy rule to identify this fault condition for the HW2000 system becomes:

Turnout type	Parameters affected by OPP-obstruction conditions									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
HW55S				↑↑	↑					↓↓
HW2000				↑	↑				↓	↓↓

Fuzzy rule to detect "OPP-obstruction" faults in HW2000 turnouts

If	R_A	is	ANY	AND	L_A	is	ANY	
AND	k_M	is	ANY	AND	J_{PM}	is	NOT Low	
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	ANY	
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY	
AND	$\omega_m(t_0)$	is	NOT High	AND	*TotDist*	is	Low	
Then	State	is	OPP-Obstruction (for the HW2000 turnout)					(6-18)

All modifications confirm that mechanical faults still only affect the mechanical turnout parameters. However, some relationships between their qualitative changes and the associated faults are no longer applicable, although the estimated parameters of the HW2000 turnout system still retaining the same physical interpretations to those of the HW55S. These differences arise because of the distinct mechanical properties between the switch rails connected to the training HW55S point machine and those connected to the HW2000. Since the switch rails in the HW2000 turnout system are much shorter than those in the HW55S, they are no longer fixed at one end, meaning that there is no spring-force resisting the motion of the switch rails. This property alters how these switch rails act as a load on the DC motor. For example, in the unlock difficulty conditions, the reduced stiffness of the rails results in less restoration force, making the unlocking process not as difficult as it was in the HW55S experiments.

Once the membership functions of all turnout parameters have been adjusted to agree with the fault-free HW2000 turnout system characteristics and all ineffective fuzzy rules have been modified, the FDD results will be recalculated in the final section.

6.4.4 Final FDD results after modification

After the fuzzy logic-based classification has been modified to be suitable for diagnosing the HW2000 turnout system, the confusion matrix is shown in Figure 6-29.

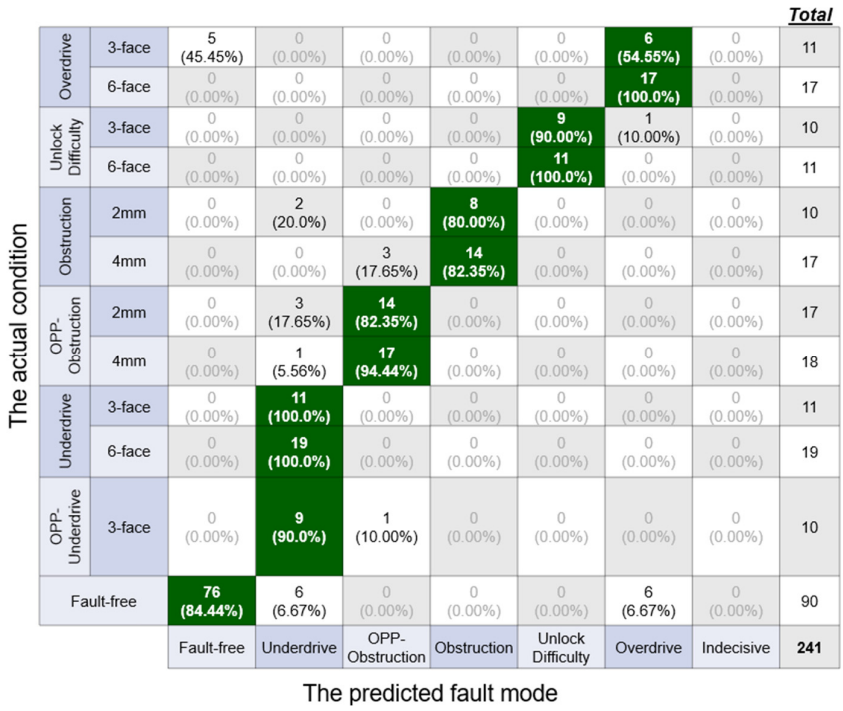


Figure 6-30 Confusion matrix showing the FDD results of the N-to-R HW2000 turnout testing dataset once the methodology has been modified to suit the HW2000 turnout

The matrix demonstrates a great improvement in the methodology's accuracy from 61.41% (before the modification) to 87.55% (after the modification). In particular, the previous misclassifications are now mostly corrected. The associated likelihood trends of the modified FDD methodology are also provided in Figure 6-30.

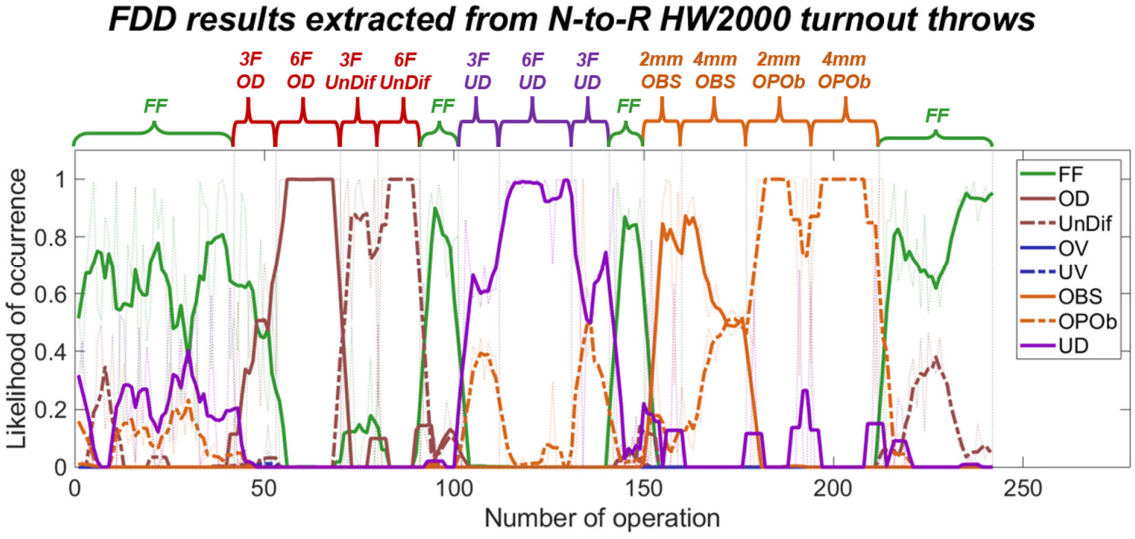


Figure 6-31 Fault occurrence likelihood trends of the entire N-to-R HW2000 testing dataset once the methodology has been modified to suit the HW2000 turnout

The results after the modification correctly identify that the turnout is more likely to suffer from unlock difficulty faults during the “6F UnDif” period. Also, the proposed FDD methodology can now accurately classify all the obstruction-affected operations, whereas the diagnoses of the others are comparable to those before the modification. The diagnostic results for all R-to-N operations are also provided in Appendix 5.

This section illustrates that a validated FDD methodology that was previously developed from one system always requires some degree of re-training before it can be effectively applied to other systems. Due to the preserved thorough interpretability and traceability of the proposed parameter estimation-based FDD methodology, the normally sophisticated re-training process can be simplified to a cross-training instead.

That is, the aspects of the methodology that still perform satisfactorily are kept unchanged, while the others are modified according to the new system characteristics. In this case, the interpretable turnout parameters can still be considered effective FDD features since both HW55S and HW2000 turnout systems use a permanent magnet DC motor as their main drive and have a similar driving mechanism. Meanwhile, the aspects that required modifications are (1) the membership functions of every turnout parameter that must be modified to match the fault-free characteristics of the HW2000 system and (2) the ineffective fuzzy rules that produced incorrect diagnostic results, including the ones to detect unlock difficulty, obstruction, and OPP-obstruction faults.

6.5 Conclusions

In this chapter, an interpretable and traceable fuzzy logic-based FDD classifier used to analyse the turnout parameter estimates was developed. The methodology’s traceability was established when the membership functions used in the fuzzification were created based on the statistical characteristics of the estimation results of the

training dataset. The interpretable fuzzy rules were also formulated from the patterns between the parameter changes and associated fault modes, provided in Chapter 5.

The diagnostic result of each operation was then displayed, including trends of occurring likelihood of every health state and the membership value of all parameters, to allow the maintainers to self-validate the diagnostic results based on their expertise. Then, the algorithm's effectiveness was validated by using an arbitrary threshold of occurrence likelihood at 0.5. It was found that the proposed methodology achieved more than 80% overall accuracy for both the HW55S training and testing datasets.

Finally, the proposed FDD methodology was also applied to a different turnout system (HW2000) to assess its transferability. Most of the same experiments carried out on the HW55S system were also conducted on the HW2000. The initial diagnostic results from the immediate use of the methodology, however, were undesirable.

Thus, the proposed FDD methodology had to undergo a re-training process first before it could diagnose the HW2000 turnout system effectively. In this case, a simpler cross-training process was enabled due to the inherent interpretability and traceability of the methodology, and the common characteristics between the two turnout systems. Hence, most elements of the methodology that were suitable for the HW55S turnout system could be immediately transferred to the HW2000, while the others had to go through a modification based on the HW2000 characteristics. Once the cross-training and tuning process had been done, the FDD performance was improved considerably for the HW2000 turnout system as the fault occurring likelihood trend now aligned with all of the actual fault conditions and the accuracy was improved from 61% to 88%.

In conclusion, this chapter developed a fuzzy logic-based FDD classifier that also aimed to preserve thorough interpretability and traceability of the methodology.

The proposed methodology performed adequately well for the training HW55S turnout system. It provided information on both the fault development trends and interpretable turnout parameters that also allowed the skilled personnel to better understand the diagnostic results. The transferability of the proposed methodology was also validated by using a different turnout system, the HW2000. It was found that the preserved interpretability and traceability of the methodology allowed a simplified cross-training process to be applied. Once the proposed FDD methodology had been cross-trained and tuned, it could also diagnose the HW2000 turnout system effectively. This implies that such an explanatory mechanism can enhance the methodology's transferability and facilitate its re-training process, improving its realistic practicality in the process.

CHAPTER 7 CONCLUSIONS AND FUTURE WORK

7.1 Conclusions of the FDD methodology development

This thesis introduces the development of a parameter estimation-based FDD methodology to diagnose the occurrence of faults in railway turnout systems. Given their critical role in changing train directions, their failure can impact both the economic and safety aspects of railway operations. Despite such significance, most systems are still currently maintained by scheduled preventive strategies, which are not financially optimal and failures may occur between inspections. Therefore, the development of FDD methodologies that can inform the current health state of a turnout system and support condition-based maintenance has become an attractive research area.

The development of FDD methodologies based on data-driven approaches has been more prevalent due to the advance of data acquisition technologies. Although previous studies have shown their exceptional diagnosis accuracy, their mathematical abstractness usually caused them to lack an explanatory mechanism. The inability to understand the diagnostic reasoning of a methodology can result in a lack of trust and confidence from the users, especially when it produces misclassifications. Besides, data-driven FDD methodologies can only provide optimal results for the system they have been trained with. When they are applied to other systems, they have to be re-trained based solely on the new training dataset that also must be comprehensive.

Due to these challenges that prevent data-driven FDD approaches from being practical for railway turnout applications, this thesis hypothesised that model-based FDD approaches that integrate underlying system knowledge into the development can be more practical alternatives if they can be demonstrated to be viable.

To evaluate the hypothesis of this thesis, four objectives were set to achieve during the course of this PhD study as follows:

1. Identify a model-based technique for developing a practical FDD methodology for railway turnouts, such that the diagnosis results should be able to identify the fault mode by an interpretable algorithm, while the proposed methodology should also be transferrable to other turnouts of the same type.
2. Assumptions and simplifications must be justified and validated by simulation studies to initially ensure the proposed model-based technique's feasibility.
3. Demonstrate the FDD capabilities of the proposed methodology by applying it to the measurements taken from an actual railway turnout system.
4. Once the performance of the proposed methodology is determined, evaluate its transferability when it is applied to another railway turnout system.

In Chapter 2, background knowledge of railway turnout systems was provided, which was followed by a literature review. This outlined a survey and a comparison of methodologies previously used in FDD developments, attempting to identify a model-based technique that was to be employed in the proposed FDD methodology. Once the requirements of a practical FDD methodology for railway turnout were considered, the parameter estimation approach was selected, achieving the first objective.

Verification of the effectiveness when the parameter estimation was used for FDD purposes was then conducted on a simulation model, as presented in Chapter 3. The transfer function models required for the estimation were formulated from first-principle knowledge of the turnout system, while its dynamics were also assumed to be linear and time-invariant to maintain the comprehensibility of the methodology.

Then, these assumptions and simplifications were validated in Chapter 4 when the same estimation algorithm and transfer function models were applied to a higher-

fidelity simulation model that represented the turnout system more realistically. The results showed that the proposed methodology can accurately estimate the turnout parameters and detect their changes when faults are injected into the simulation model. Thus, these two chapters have validated the initial feasibility of the parameter estimation technique with simulation models, corresponding to the second objective.

The third objective of implementing the proposed methodology to estimate the parameters of a real HW55S turnout system was achieved in Chapter 5. Due to the differences between previous simulation models and the real HW55S turnout system, a segmentation process was incorporated. This would divide any turnout operation which was realistically time-variant and non-linear into two linear phases, allowing the parameter estimation algorithm to be successfully applied to each phase at a time. Five fault modes were then simulated onto the HW55S turnout system and parameter estimates from their measurements were obtained. By comparing them to the baseline estimates when the turnout was fault-free, unique relationships between the qualitative changes in parameters and associated fault modes were determined.

In Chapter 6, a fuzzy logic-based FDD algorithm was developed to evaluate the turnout parameters, by using interpretable and approximated human reasoning. The proposed FDD methodology achieved more than 80% overall accuracy when it was applied to diagnose the training HW55S turnout system. Although it seemed to be inferior compared with other data-driven methods, the use of interpretable parameters in the diagnosis would allow skilled personnel to re-assess every diagnostic result. This also came along with the occurrence likelihood trend of each health state, which would enable maintenance staff to be notified of ongoing fault development over time.

Finally, the transferability of the proposed FDD methodology when applied to a different turnout system, HW2000, was also validated in Chapter 6, in order to achieve

the last objective. It was found that the interpretable turnout parameters and intuitive fuzzy logic reasoning used in the classification algorithm facilitated the application of the cross-training process, which is much less sophisticated than the usual re-training procedure. This improved the diagnosis of the HW2000 turnout system drastically to 87.55% accuracy, compared to 61.41% before the tuning and cross-training.

From all four objectives that were achieved during the study, the hypothesis is re-stated here before the final evaluation:

"By incorporating physics knowledge of railway turnouts into model-based FDD development, a more transferable methodology that can provide sufficiently accurate, comprehensive, and interpretable diagnosis results can be established",

It was found that the application of parameter estimation which incorporated physics knowledge of turnout systems into the FDD methodology development was feasible. The proposed methodology achieved more than 80% accuracy with the ability of multiple fault classification. Moreover, the diagnostic reasoning could be developed in a comprehensible and traceable manner, since the physically interpretable turnout parameters were used in the evaluation. This can facilitate the application of cross-training and tuning procedure as well, which can greatly simplify the re-training process when the proposed methodology is to be used to diagnose other turnout systems.

7.2 Recommendations and future works

In this thesis, since it aimed to validate the feasibility of the model-based FDD methodology when applied to a working railway turnout system, the fault simulation process was kept simple by simulating only one fault mode on the turnout at a time. This was done to ensure that there were no overlapping effects.

Nevertheless, this simplification certainly did not cover the entire range of possibilities, since realistically, multiple fault conditions may occur on the turnout at the same time. As a result, the first recommendation is to further investigate the effects on turnout parameters when multiple fault modes occur concurrently and identify an approach that can classify their overlapping effects effectively.

The second recommendation focuses on the actual implementation practice. As already demonstrated with a system overview flowchart in Figure 6-27, to implement this FDD methodology for real applications, first, obtaining the fault-free baseline value of every turnout parameter is compulsory. Such fault-free experiments can be done after installation or maintenance. Doing so would allow influences from the surrounding environment to affect the monitored turnout system as well, making all baseline values more representative. This process must be done individually as each turnout has different characteristics. As for the fuzzy rules, if the turnout systems are operated with the same working principle, the relationships between parameter changes and associated fault modes can be transferrable. This knowledge could be obtained via a few fault simulations conducted on a test turnout system.

The last recommendation emphasises that the practicality of the proposed FDD methodology has only been demonstrated for HW turnouts. There are other commonly used electro-mechanical turnout models that operate with similar working principles, such as M63, P80, and HPSS. In future studies, if this methodology can be effectively applied to those turnouts with a few reconfigurations, it will validate the generalisability and widen the practicality of this FDD methodology as well. Combined with the use of only three sensors to obtain measurements for effective fault diagnosis while also being notified of fault development trends, this will greatly enhance the possibility of industrial implementation to facilitate condition-based maintenance in the future.

REFERENCES AND BIBLIOGRAPHY

Alfi, S., Bionda, S., Bruni, S. & Gasparetto, L., 2011. Condition monitoring of suspension components in railway bogies. *5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing (RCM 2011),* pp. 1-6.

Antong, H., Dixon, R. & Ward, C. P., 2017. Least squares estimation and fuzzy logic based condition monitoring for high redundancy actuator. *IFAC-PapersOnLine,* 50(1), pp. 3189-3194.

Ardakani, H. D. et al., 2012. PHM for railway system — A case study on the health assessment of the point machines. *2012 IEEE Conference on Prognostics and Health Management,* pp. 1-5.

Asada, T., 2013. *Novel condition monitoring techniques applied to improve the dependability of railway point machines (PhD Thesis).* Birmingham: University of Birmingham.

Asada, T. & Roberts, C., 2011. Development of an effective condition monitoring system for AC point machines. *5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing (RCM 2011),* pp. 1-6.

Asada, T. & Roberts, C., 2013. Improving the dependability of DC point machines with a novel condition monitoring system. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit,* 227(4), pp. 322-332.

Asada, T., Roberts, C. & Koseki, T., 2013. *An algorithm for improved performance of railway condition monitoring equipment: Alternating-current point machine case study.* s.l., s.n., pp. 81-92.

Asadi, S. & Shahrabi, J., 2016. RipMC: RIPPER for multiclass classification. *Neurocomputing,* Volume 191, pp. 19-33.

Åström, K. J. & Eykhoff, P., 1971. System identification—A survey. *Automatica,* 7(2), pp. 123-162.

Atamuradov, V. & Camci, F., 2017. Segmentation based feature evaluation and fusion for prognostics. *International Journal of Prognostics and Health Management,* 8(2), pp. 1-14.

Atamuradov, V., Camci, F., Baskan, S. & Sevkli, M., 2009. *Failure Diagnostics for Railway Point Machines Using Expert Systems.* s.l., IEEE, pp. 1-5.

Atamuradov, V. et al., 2018a. Railway point machine prognostics based on feature fusion and health state assessment. *IEEE Transactions on Instrumentation and Measurement,* pp. 2691-2704.

Atamuradov, V. et al., 2018b. Degradation-level assessment and online prognostics for sliding chair failure on point machines. *IFAC-PapersOnLine,* 51(24), pp. 208-213.

Atamuradov, V. et al., 2019. Feature selection and fault-severity classification–based machine health assessment methodology for point machine sliding-chair degradation. *Quality and Reliability Engineering International,* 35(4), pp. 1081-1099.

Atamuradov, V. et al., 2020. Machine health indicator construction framework for failure diagnostics and prognostics. *Journal of signal processing systems,* Volume 92, pp. 591-609.

Bagwell, P., 2004. The Sad State of British Railways: The Rise and Fall of Railtrack, 1992–2002. *The Journal of Transport History,* 25(2), pp. 111-124.

Bai, H., 2010. *A generic fault detection and diagnosis approach for pneumatic and electric driven railway assets (PhD Thesis).* Birmingham: University of Birmingham.

Bell, C., 2005. Point, track circuit and level crossing condition monitoring. *IEE Seminar on Railway Condition Monitoring, 2005 (Ref. No. 2005/10813),* pp. 5-12.

Bemment, S. D. et al., 2017. Rethinking rail track switches for fault tolerance and enhanced performance. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 231(9), pp. 1048-1065.

Bemment, S. D., Goodall, R. M., Dixon, R. & Ward, C. P., 2018. Improving the reliability and availability of railway track switching by analysing historical failure data and introducing functionally redundant subsystems. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit,* pp. 1407-1424.

Bian, C. et al., 2018. Degradation detection method for railway point machines. pp. 1-25.

Bian, C. et al., 2019. Degradation state mining and identification for railway point machines. *Reliability Engineering & System Safety,* Volume 188, pp. 432-443.

Böhm, T., 2017. Remaining useful life prediction for railway switch engines using classification techniques. *International Journal of Prognostics and Health Management,* Volume 8, pp. 1-15.

Bradbeer, T., 2022. Back to basics: Points maintenance and remote condition monitoring. *IRSE News*, April, pp. 24-29.

British Railways Board, 1992. *Electrical Detection and Power Operated Points,* s.l.: s.n.

Bukhsh, Z. A., Saeed, A., Stipanovic, I. & Doree, A. G., 2019. *Predictive maintenance using tree-based classification techniques: A case of railway switches.* s.l., s.n., pp. 35-54.

Bunnett, R., 2015. *Alstom HW1000/2000 point machines.* [Online] Available at: https://www.youtube.com/watch?v=pmsMpEszQsA [Accessed 30 August 2022].

Camci, F., Eker, O. F., Başkan, S. & Konur, S., 2016. Comparison of sensors and methodologies for effective prognostics on railway turnout systems. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 230(1), pp. 24-42.

Cao, Y., Sun, Y., Li, P. & Su, S., 2023. Vibration-Based Fault Diagnosis for Railway Point Machines Using Multi-Domain Features, Ensemble Feature Selection and SVM. *IEEE Transactions on Vehicular Technology,* pp. 1-9.

Cao, Y., Sun, Y., Xie, G. & Li, P., 2021. A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier. *IEEE Transactions on Intelligent Transportation Systems,* pp. 1-10.

Chamroukhi, F., Same, A., Aknin, P. & Antoni, M., 2008. Switch mechanism diagnosis using a pattern recognition approach. *4th IET International Conference on Railway Condition Monitoring,* p. 1-4.

Cheng, Y. & Zhao, H., 2015. *Fault detection and diagnosis for railway switching points using fuzzy neural network.* s.l., IEEE, pp. 860-865.

Chen, J. & Roberts, C., 2006. Effective Condition Monitoring of Line Side Assets. *2006 IET International Conference On Railway Condition Monitoring,* pp. 78-83.

Chen, Q. et al., 2021. Improved Fault Diagnosis of Railway Switch System Using Energy-Based Thresholding Wavelets (EBTW) and Neural Networks. *IEEE Transactions on Instrumentation and Measurement,* Volume 70, pp. 1-12.

Chen, Q., Nicholson, G., Ye, J. & Roberts, C., 2020. *Fault Diagnosis Using Discrete Wavelet Transform (DWT) and Artificial Neural Network (ANN) for A Railway Switch.* s.l., s.n., pp. 67-71.

Chow, E. & Willsky, A., 1984. Analytical redundancy and the design of robust failure detection systems. *IEEE Transactions on automatic control,* 29(7), pp. 603-614.

Cui, K., Tang, M. & Ou, D., 2019. Simulation data generating algorithm for railway turnout fault diagnosis in big data maintenance management system. *4th International Symposium for Intelligent Transportation and Smart City, ITASC 2019*, pp. 155-166.

Dassanayake, H., Roberts, C., Goodman, C. J. & Tobias, A. M., 2009. Use of parameter estimation for the detection and diagnosis of faults on electric train door systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* 223(4), pp. 271-278.

Department of Rail Transport, 2022. *Rail Fleet Passenger Statistics during a new emergence of COVID-19 (translated from ปริมาณสถิติผู้โดยสารในระบบขนส่งทางรางทุกระบบ ณ สถานการณ์การแพร่ระบาด COVID-19 ระลอกใหม่).* [Online] Available at: https://www.drt.go.th/library/passengerstat [Accessed 28 September 2022].

Department of Rail Transport, 2023. *Statistics of Rail Passenger (translated from ข้อมูลสถิติผู้โดยสารรถไฟฟ้าในเมือง).* [Online] Available at: https://drt.gdcatalog.go.th/dataset/stat_pass_rail [Accessed 23 July 2023].

Dilys, J. & Stankevič, V., 2022. A Simple Method for Stator Inductance and Resistance Estimation for PMSM at Standstill. *International Journal of Robotics and Control Systems,* 2(3), pp. 477-491.

Ding, S. X., 2008. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools.* s.l.:Springer Science & Business Media.

Dixon, R., 2000. Health monitoring of an electrical actuation system: A parameter estimation based approach. *Proceedings of 14th International Conference on Systems Engineering, ICSE2000, Conventry University,* Volume 1, pp. 123-128.

Dixon, R. & Pike, A. W., 2002. Application of condition monitoring to an electromechanical actuator: A parameter estimation based approach. *Computing and Control Engineering Journal,* 13(2), pp. 71-81.

Dowdeswell, B., Sinha, R. & MacDonell, S. G., 2020. Finding faults: A scoping study of fault diagnostics for Industrial Cyber–Physical Systems. *Journal of Systems and Software,* Volume 168, pp. 1-16.

Du, H. et al., 2019. Dynamic Time Warping and Spectral Clustering Based Fault Detection and Diagnosis of Railway Point Machines. *2019 IEEE Intelligent Transportation Systems Conference (ITSC),* pp. 595-600.

Dutta, S. et al., 2020b. Health monitoring of a railway track switch actuator based on continuous-time parameter estimation method. *8th Transport Research Arena TRA 2020,* pp. 1-9.

Dutta, S. et al., 2022. Continuous Time Parameter Estimation Method for a Railway Track Switch Actuator. *2022 UKACC 13th International Conference on Control (CONTROL),* pp. 56-59.

Dutta, S. et al., 2018. Modelling and controller design for self-adjusting railway track switch system. *VIENNA 2018 A digital era for transport ,* pp. 1-10.

Dutta, S. et al., 2023. A Framework for Dynamic Modelling of Railway Track Switches Considering the Switch Blades, Actuators and Control Systems. *Railway Engineering Science,* pp. 1-19.

Dutta, S. et al., 2020a. A new approach to railway track switch actuation: Dynamic simulation and control of a self-adjusting switch. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit v.234,* pp. 779-790.

Eker, O., Camci, F. & Kumar, U., 2012. SVM based diagnostics on railway turnouts. *International Journal of Performability Engineering,* 8(3), pp. 289-298.

Eker, O. F. et al., 2011. A simple state-based prognostic model for railway turnout systems. *IEEE Transactions on Industrial Electronics,* 58(5), pp. 1718-1726.

Fidali, M., Wojciechowski, P. & Pełka, A., 2018. Fault Detection of Railway Point Machine Using Diagnostic Models. *Advances in Technical Diagnostics: Proceedings of the 6th International Congress on Technical Diagnostic, ICDT2016, 12-16 September 2016, Gliwice, Poland 6,* pp. 275-285.

Forrai, A., 2017. System Identification and Fault Diagnosis of an Electromagnetic Actuator. *IEEE Transactions on Control Systems Technology,* 25(3), pp. 1028-1035.

Frank, P. M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica,* 26(3), pp. 459-474.

Gao, Y., Yang, Y., Ma, Y. & Xu, W., 2022. Study on Intelligent Diagnosis of Railway Turnout Switch Based on Improved FastDTW and Time Series Segmentation under Big Data Monitoring. *Mathematical Problems in Engineering.*

Gao, Z., Cecati, C. & Ding, S. X., 2015a. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE transactions on industrial electronics,* 62(6), pp. 3757-3767.

Garnier, H., 2015. Direct continuous-time approaches to system identification. Overview and benefits for practical applications. *European Journal of Control,* Volume 24, pp. 50-62.

Garnier, H., Mensler, M. & Richard, A., 2003. Continuous-time model identification from sampled data: Implementation issues and performance evaluation. *International Journal of Control,* 76(13), pp. 1337-1357.

Garnier, H. & Wang, L., 2008. *Identification of continuous-time models from sampled data.* Springer.

Gilson, M., 2015. What has Instrumental Variable method to offer for system identification?. *IFAC-PapersOnLine,* 48(1), pp. 354-359.

Goodall, R. M. & Roberts, C., 2006. Concepts and Techniques for Railway Condition Monitoring. *2006 IET International Conference On Railway Condition Monitoring,* pp. 90-95.

Goodwin, M. J., 1998. Recent Methods and Trends in Determining Point Maintenance Frequencies. *Proceedings of the Institution of Civil Engineers - Transport,* 129(2), pp. 102-105.

Guangwu, C. & Rong, L., 2021. Method of Turnout Fault Diagnosis Based on DBN-BiLSTM Model. *2021 CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes (SAFEPROCESS),* pp. 1-6.

Guillaume, S. & Charnomordic, B., 2011. Learning interpretable fuzzy inference systems with FisPro. *Information Sciences,* 181(20), pp. 4409-4427.

Guo, Z., Wan & Y. Y. H., 2020b. An Unsupervised Fault-Detection Method for Railway Turnouts. *IEEE Transactions on Instrumentation and Measurement,* 69(11), pp. 8881-8901.

Guo, Z., Ye, H., Jiang, M. & Sun, X., 2020a. An enhanced fault detection method for railway turnouts incorporating prior faulty information. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC),* pp. 1-6.

Guzman, D. N. et al., 2019. Anomaly detection and forecasting methods applied to point machine monitoring data for prevention of switch failures. *COMADEM 2019 Proceedings,* pp. 1-11.

Guzman, D. N. et al., 2020. Detecting anomalous behavior of railway switches under real operation conditions: Workflow and implementation. *TRA 2020 conference proceedings,* pp. 1-7.

Guzman, D. N. et al., 2018. *Data-driven condition now-and forecasting of railway switches for improvement in the quality of railway transportation.* s.l., Proceedings of the European Conference of the PHM Society.

Guzman, D. N. et al., 2022. *Towards the automation of anomaly detection and integrated fault identification for railway switches in a real operational environment.* Birmingham, s.n.

Hamadache, M. et al., 2019b. *Residual-based fault detection method: application to railway switch & crossing (S&C) system.* s.l., IEEE, pp. 1228-1233.

Hamadache, M. et al., 2019a. On the Fault Detection and Diagnosis of Railway Switch and Crossing Systems: An Overview. *Applied Sciences v.9,* pp. 5129-5160.

Hao, B., Roberts, C. & Goodman, C., 2008. A generic fault detection and diagnosis approach for railway assets. *2008 International Conference on Railway Engineering - Challenges for Railway Transportation in Information Age,* pp. 1-8.

Hassankiadeh, J. S., 2011. *Failure analysis of railway switches and crossings for the purpose of preventive maintenance (MSc Dissertation).* Stockholm : Royal Institute of Technology.

How, F., 2022a. Back to basics: Points Part 1. *IRSE News 285*, February, pp. 13-21.

How, F., 2022b. Back to basics: Points Part 2. *IRSE News*, March, pp. 11-17.

Huang, S. et al., 2017. Turnout fault diagnosis through dynamic time warping and signal normalization. *Journal of Advanced Transportation,* pp. 1-8.

Hu, X., Cao, Y., Tang, T. & Sun, Y., 2022. Data-driven technology of fault diagnosis in railway point machines: Review and challenges. *Transportation Safety and Environment,* 4(4), pp. 1-14.

Hu, X., Tang, T., Tan, L. & Zhang, H., 2023. Fault Detection for Point Machines: A Review, Challenges, and Perspectives. *Actuators,* 12(10), pp. 1-21.

IAD Rail Systems, 2001. *High Performance Switch System powered by the High Performance Switch Actuator and the 'POWERLINK' Backdrive,* s.l.: s.n.

IFAC, 2019. *Terminology in the Area of Fault Management.* [Online] Available at: https://tc.ifac-control.org/6/4/terminology/terminology-in-the-area-of-fault-management [Accessed 20 August 2022].

Igarashi, Y. & Siomi, S., 2006. Development of monitoring system for electric switch machine. *Quarterly Report of RTRI,* 47(2), pp. 78-82.

Independent Regulations' Group - Rail, 2017. *Overview on European Performance Schemes.* [Online] Available at: https://www.irg-rail.eu/download/5/948/IRG-Rail20227-OverviewoftheImplementationofDirectCostsinEurope.pdf [Accessed 20 April 2023].

INNOTRACK, 2008a. *D3.3.1 - List of key parameters for switch and crossing monitoring.* [Online] Available at: https://www.charmec.chalmers.se/innotrack/deliverables/sp3/d331-f3p-key_parameters_switch_crossing_monitoring.pdf

INNOTRACK, 2008b. *D3.3.2 - Available sensors for railway environments for condition monitoring.* [Online] Available at: https://www.charmec.chalmers.se/innotrack/deliverables/sp3/d332-f3p-available_sensors.pdf

Isermann, R., 2005. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control,* 29(1), pp. 71-85.

Isermann, R., 2006. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance.* s.l.:Springer Science & Business Media.

Isermann, R., 2011. *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems.* s.l.:Springer Science & Business Media.

Isermann, R. & Balle, P., 1997. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control engineering practice,* 5(5), pp. 709-719.

ISO, 2003. *ISO 13372:2003 - Condition monitoring and diagnostics of machines.* s.l.:s.n.

Iwasawa, N., Ryuo, S., Kawasaki, K. & Hada, A., 2015. Development of System for Supporting Lock Position Adjustment Work for Electric Point Machine. *Quarterly Report of RTRI,* 56(3), pp. 200-205.

Jackson, M., 2023. *HPSS Mk2,* Glasgow: The PWI.

Jackson, M. & Spurlock, S., 2021. *Points Operating Equipment & High Performance Switch System (HPSS) - Mark Jackson & Steve Spurlock.* [Online] Available at: https://www.youtube.com/watch?v=FPtGOIzmeY4&t=1054s [Accessed 10 November 2022].

Jafarian, E. & Rezvani, M. A., 2012. Application of fuzzy fault tree analysis for evaluation of railway safety risks: an evaluation of root causes for passenger train derailment. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 226(1), pp. 14-25.

Jakeman, A. J., Steele, L. P. & Young, P. C., 1980. Instrumental Variable Algorithms for Multiple Input Systems Described by Multiple Transfer Functions. *IEEE Transactions on Systems, Man, and Cybernetics,* 10(10), pp. 593-602.

Janot, A., Young, P. C. & Gautier, M., 2017. Identification and control of electro-mechanical systems using state-dependent parameter estimation. *International Journal of Control,* 90(4), pp. 643-662.

Jiang, Y. et al., 2017. Data-based fractional differential models for non-linear dynamic modeling of a lithium-ion battery. *Energy,* Volume 135, pp. 171-181.

Jimenez, J. J. M. et al., 2020. Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems,* Volume 56, pp. 539-557.

Jin, W. et al., 2015. Development and evaluation of health monitoring techniques for railway point machines. *2015 IEEE Conference on Prognostics and Health Management (PHM),* pp. 1-11.

Kim, H. et al., 2016. Fault diagnosis of railway point machines using dynamic time warping. *Electronics Letters,* 52(10), pp. 818-819.

Kim, S. H. & Tahk, M.-J., 2018. Dynamic Stiffness Transfer Function of an Electromechanical Actuator Using System Identification. *International Journal of Aeronautical and Space Sciences,* 19(1), pp. 208-216.

Kumpao, T., 2019. *An Algorithm Development of Condition Monitoring System for Point Machine; MSc Dissertation.* Birmingham: University of Birmingham.

Kun, S. Y., Tao, H. & Cheng, W. Y., 2021. Research on Fault Diagnosis Method of Switch Machine Based on KFCM. *Journal of Physics: Conference Series,* 1972(1), pp. 1-9.

Kwon, Y. W. & Bang, H., 2000. *The Finite Element Method Using MATLAB, Second Edition.* 2nd ed. s.l.:CRC Press.

Ladj, A., Bouamama, B. O. & Toguyeni, A., 2022. Comparison Between Model-Based and Data-Based Methods for Fault Diagnosis of Railway Turnouts. *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT),* Volume 1, pp. 617-622.

Lee, J. et al., 2016. Fault detection and diagnosis of railway point machines by sound analysis. *Sensors,* 16(4), pp. 1-12.

Letot, C. et al., 2015. A data driven degradation-based model for the maintenance of turnouts: a case study. *IFAC-PapersOnLine,* 48(21), pp. 958-963.

Li, L., Dutta, S., Dixon, R. & Stewart, E., 2022. Fault Tolerant Actuation of a Railway Track Switch: a Simulation Study. *2022 UKACC 13th International Conference on Control (CONTROL),* pp. 207-212.

Li, L., Dutta, S., Dixon, R. & Stewart, E., 2023. Railway track switch simulation: a new dynamic model for studying actuator and switch blade dynamics. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 237(6), pp. 775-783.

Li, M. & Fei, R., 2022. Research and Implementation of Fault Diagnosis of Switch Machine Based on Data Enhancement and CNN. *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom),* pp. 1467-1472.

Li, M. et al., 2022. A Fault-Diagnosis Method for Railway Turnout Systems Based on Improved Autoencoder and Data Augmentation. *Sensors,* 22(23), pp. 1-16.

Ling, X. & Haldar, A., 2004. Element level system identification with unknown input with Rayleigh damping. *Journal of Engineering Mechanics,* 130(8), pp. 877-885.

Liu, J., Wen, T., Xie, G. & Cao, Y., 2022. Modified Multi-Scale Symbolic Dynamic Entropy and Fuzzy Broad Learning-Based fast fault diagnosis of Railway Point Machines. *Transportation Safety and Environment,* pp. 1-8.

Li, Z., Dai, S., Zheng, Z. & Liang, Y., 2021. Fault diagnosis of ZD6 turnout system based on wavelet transform and GAPSO-FCM. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC),* pp. 2187-2192.

Li, Z., Yin, Z., Tang, T. & Gao, C., 2019. Fault Diagnosis of Railway Point Machines Using the Locally Connected Autoencoder. *Applied Sciences,* 9(23), pp. 1-17.

Ljung, L., 1999. *System identification: Theory for the User.* 2nd ed. New Jersey: Prentice-Hall, Inc..

Márquez, F. P. G., 2010. A new method for maintenance management employing principal component analysis. *Structural Durability & Health Monitoring,* 6(2), pp. 89-99.

Márquez, F. P. G., Lewis, R. W., Tobias, A. M. & Roberts, C., 2008. *Life cycle costs for railway condition monitoring.* s.l., s.n., pp. 1175-1187.

Márquez, F. P. G. & Pedregal, D. J., 2006. An Algorithm for Detecting Faults in Railway Point Mechanisms. *IFAC Proceedings Volumes,* 39(13), pp. 1360-1365.

Márquez, F. P. G., Roberts, C. & Tobias, A. M., 2010. Railway point mechanisms: Condition monitoring and fault detection. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 224(1), pp. 35-44.

Márquez, F. P. G. & Schmid, F., 2007. *A digital filter-based approach to the remote condition monitoring of railway turnouts.* s.l., s.n., pp. 830-840.

Márquez, F. P. G., Schmid, F. & Collado, J. C., 2003a. *A reliability centered approach to remote condition monitoring. A railway points case study.* s.l., s.n., pp. 33-40.

Márquez, F. P. G., Schmid, F. & Collado, J. C., 2003b. Wear assessment employing remote condition monitoring: a case study. *Wear,* 225(7-12), pp. 1209-1220.

Márquez, F. P. G., Tercero, D. J. P. & Schmid, F., 2007. Unobserved Component models applied to the assessment of wear in railway points: A case study. *European Journal of Operational Research,* 176(3), pp. 1703-1712.

Márquez, F. P. G., Weston, P. & Roberts, C., 2007. *Failure analysis and diagnostics for railway trackside equipment.* s.l., s.n., pp. 1411-1426.

Maruta, I. & Sugie, T., 2013. Projection-based identification algorithm for grey-box continuous-time models. *Systems & Control Letters,* 62(11), pp. 1090-1097.

Matei, I., Ganguli, A., Honda, T. & de Kleer, J., 2015. The Case for a Hybrid Approach to Diagnosis: A Railway Switch. *DX@ Safeprocess*, pp. 225-234.

Mathwork®, 2023. *System Identification Toolbox™ - User Guide.* [Online] Available at: https://www.mathworks.com/help/pdf_doc/ident/ident_ug.pdf [Accessed 11 November 2023].

Mathworks®, 2023. *Fuzzy Logic Toolbox™ - User's Guide.* [Online] Available at: https://www.mathworks.com/help/pdf_doc/fuzzy/fuzzy_ug.pdf [Accessed 20 January 2024].

McHutchon, M. A., Staszewski, W. J. & Schmid, F., 2005. Signal processing for remote condition monitoring of railway points. *Strain 41,* 41(2), pp. 71-85.

Miller, S., Soares, T., Weddingen, Y. V. & Wendlandt, J., 2017. *Modeling Flexible Bodies with Simscape Multibody Software - An Overview of Two Methods for Capturing the Effects of Small Elastic Deformations,* s.l.: Mathworks®.

Mistry, P., Lane, P. & Allen, P., 2020. Railway point-operating machine fault detection using unlabeled signaling sensor data. *Sensors,* p. 2692.

Mistry, P. et al., 2019. Condition monitoring of motorised devices for smart infrastructure capabilities. *Smart City and Informatization: 7th International Conference, iSCI 2019, Guangzhou, China, November 12–15, 2019, Proceedings 7,* pp. 392-403.

Narges, K. H., Ahmad, M. & Fereydoun, G. M., 2022. A hybrid fault diagnosis scheme for railway point machines by motor current signal analysis. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 236(9), pp. 1026-1034.

Olaby, O. et al., 2021. Realisation of a Novel Functionally Redundant Actuation System for a Railway Track-Switch. *Applied Sciences,* 11(2), pp. 1-20.

ORR, 2021. *Rail Statistics Compendium 2020-21 Annual.* [Online] Available at: https://dataportal.orr.gov.uk/media/2038/rail-statistics-compendium-2020-21.pdf [Accessed 8 January 2023].

ORR, 2023. *Passenger rail usage January to March 2023.* [Online] Available at: https://dataportal.orr.gov.uk/media/2207/passenger-rail-usage-jan-mar-2023.pdf [Accessed 20 June 2023].

Ou, D., Ji, Y., Zhang, R. & Liu, H., 2020. An online classification method for fault diagnosis of railway turnouts. *Sensors (Switzerland),* 20(16), pp. 1-16.

Ou, D., Tang, M., Xue, R. & Yao, H., 2018. Hybrid fault diagnosis of railway switches based on the segmentation of monitoring curves. *Eksploatacja i Niezawodność,* 20(4), pp. 514-522.

Ou, D. & Xue, R., 2018. Railway Turnout Fault Analysis Based on Monitoring Data of Point Machines. *2018 Prognostics and System Health Management Conference (PHM-Chongqing),* pp. 546-552.

Ou, D., Xue, R. & Cui, K., 2019. A data-driven fault diagnosis method for railway turnouts. *Transportation research record,* 2673(4), pp. 448-457.

Overschee, P. V. & De Moor, B., 2012. *Subspace identification for linear systems: Theory—Implementation—Applications.* s.l.:Springer Science & Business Media.

Oyebande, B. & Renfrew, A., 2002. Condition monitoring of railway electric point machines. *IEE Proceedings-Electric Power Applications,* 149(6), pp. 465-473.

Oyebande, B. & Renfrew, A., 2007. The design and implementation of a simulated trackside environment test system for an electric point machine. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 221(4),* pp. 477-485.

Padilla, A., 2017. *Recursive identification of continuous-time systems with time-varying parameters (PhD Thesis).* Nancy: Université de Lorraine.

Padilla, A. et al., 2019. Identification of continuous-time models with slowly time-varying parameters. *Control Engineering Practice,* Volume 93, pp. 1-12.

Panchenko, S. et al., 2019. Critical failures of turnouts: expert approach. *Procedia Computer Science,* Volume 149, pp. 422-429.

Pan, S., González, R. A., Welsh, J. S. & Rojas, C. R., 2020a. Consistency analysis of the Simplified Refined Instrumental Variable method for Continuous-time systems. *Automatica,* Volume 113, p. 1-12.

Pan, S., Welsh, J. S., González, R. A. & Rojas, C. R., 2020b. Efficiency analysis of the Simplified Refined Instrumental Variable method for Continuous-time systems. *Automatica,* Volume 121, p. 1-10.

Panzer, H., Hubele, J., Eid, R. & Lohmann, B., 2009. *Generating a parametric finite element model of a 3d cantilever timoshenko beam using matlab.* s.l., Lehrstuhl für Regelungstechnik.

Parkinson, R., 2019. From Data to Diagnosis: Point condition monitoring through machine learning. *From Data to Diagnosis: Point condition monitoring through machine learning,* pp. 1-12.

Pascu, V., Garnier, H., Ljung, L. & Ljung, L., 2019. Benchmark problems for continuous-time model identification: Design aspects, results and perspectives. *Automatica,* Volume 107, pp. 511-517.

Patcharaprakiti, N. et al., 2010. Modeling of Single Phase Inverter of Photovoltaic System Using System Identification. *2010 Second International Conference on Computer and Network Technology,* pp. 462-466.

Pecht, M. & Jaai, R., 2010. A prognostics and health management roadmap for information and electronics-rich systems. *Microelectronics Reliability v.50,* pp. 317-323.

Pedregal, D. J., Garcia, F. P. & Roberts, C., 2009. *An algorithmic approach for maintenance management based on advanced state space systems and harmonic regressions.*, pp. 109-124.

Pratama, Z. A. & Hidayat, F., 2022. *Predictive Maintenance on Railway Turnout System: A Systematic Literature Review.* s.l., IEEE.

Raghav, P., 2018. *Understanding of Convolutional Neural Network (CNN) — Deep Learning.* [Online] Available at: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148 [Accessed 20 October 2023].

Rail Accident Investigation Branch, 2011. *Rail Accident Report - Derailment at Grayrigg 23 February 2007.* [Online] Available at: https://assets.publishing.service.gov.uk/media/547c9037ed915d4c0d000199/R202008_081023_Grayrigg_v5.pdf [Accessed 20 September 2023].

Rao, G. P. & Unbehauen, H., 2006. Identification of continuous-time systems. *IEE Proceedings-Control theory and applications,* 153(2), pp. 185-220.

Roberts, C., Dassanayake, H. P. B., Lehtasab, N. & Goodman, C. J., 2002. *Distributed quantitative and qualitative fault diagnosis: railway junction case study.* s.l., s.n., pp. 419-429.

Romero, D. F. Z., Vélez, A. S. & Gómez-Mendoza, J. B., 2019. Non-linear Grey-Box Models Applied to DC Motor Identification. *2019 IEEE 4th Colombian Conference on Automatic Control (CCAC),* p.1-5.

Sa, J. et al., 2017a. Replacement condition detection of railway point machines using an electric current sensor. *Sensors v.17,* p. 263.

Sa, J. et al., 2017b. Aging Detection of Electrical Point Machines Based on Support Vector Data Description. *Symmetry,* 9(12), pp. 1-11.

Sarmiento-Carnevali, M. et al., 2017. Design, construction, deployment and testing of a full-scale Repoint Light track switch (I).

SCI Verkehr GmbH, 2021. *The European railway market 2021 - COVID-19 crisis and market trends.* [Online] Available at: https://www.sci.de/fileadmin/user_upload/free-downloads/SCI_Verkehr_Europaeischer_Eisenbahnmarkt_2021_.pdf [Accessed 10 June 2023].

S-CODE, 2019. *D3.1 Report – Next generation control: monitoring and sensor systems.* [Online] Available at: http://www.s-code.info/media/1100/d3-1-next-generation-control-monitoring-and-sensor-systems.pdf [Accessed 20 March 2023].

Shaw, D. C., 2008. A universal approach to Points Condition Monitoring. *2008 4th IET International Conference on Railway Condition Monitoring,* pp. 1-6.

Shi, Z., Du, Y. & Yao, X., 2023. Fault diagnosis of ZDJ7 railway point machine based on improved DCNN and SVDD classification. *IET Intelligent Transport Systems.*

Shi, Z., Liu, Z. & Lee, J., 2018. An auto-associative residual based approach for railway point system fault detection and diagnosis. *Measurement,* Volume 119, pp. 246-258.

Siemens, 2014. *Operating and Service Manual - Switchguard S700K Point Machine,* s.l.: Siemens.

Silmon, J. A., 2009. *Operational industrial fault detection and diagnosis: railway actuator case studies (PhD Thesis).* Birmingham: University of Birmingham.

Silmon, J. & Roberts, C., 2010. *Improving railway switch system reliability with innovative condition monitoring algorithms.* s.l., s.n., pp. 293-302.

Silmon-Monerri, J. & Roberts, C., 2006. A Systems Approach to Fault Detection and Diagnosis for Condition-Based Maintenance. *2006 IET International Conference On Railway Condition Monitoring,* pp. 141-148.

Soares, N., Aguiar, E. P. d., Souza, A. C. & Goliatt, L., 2021. Unsupervised machine learning techniques to prevent faults in railroad switch machines. *International Journal of Critical Infrastructure Protection,* Volume 33.

Söderström, T., 2011. A generalized instrumental variable estimation method for errors-in-variables identification problems. *Automatica,* 47(8), pp. 1656-1666.

Söderström, T. & Stoica, P., 1985. Instrumental Variable Methods for Identification of Dynamic Systems. *IFAC Proceedings Volumes,* 18(5), pp. 17-28.

Söderström, T. & Stoica, P., 2002. Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing,* 21(1), pp. 1-9.

State Railway of Thailand, 2019. *122 Years and The Future of SRT.* [Online] Available at: https://www.railway.co.th/RailwayMiddleFile/NewsAndActImg/1424/132055630040840168_SRT-122years.pdf [Accessed 20 June 2023].

Sun, Y., Cao, Y. & Li, P., 2022. Contactless fault diagnosis for railway point machines based on multi-scale fractional wavelet packet energy entropy and synchronous optimization strategy. *IEEE Transactions on Vehicular Technology,* 71(6), pp. 5906-5914.

Sun, Y., Cao, Y., Xie, G. & Wen, T., 2020. Condition monitoring for railway point machines based on sound analysis and support vector machine. *Chinese Journal of Electronics,* 29(4), pp. 786-792.

Sun, Y., Cao, Y., Xie, G. & Wen, T., 2021. Sound based fault diagnosis for RPMs based on multi-scale fractional permutation entropy and two-scale algorithm. *IEEE Transactions on Vehicular Technology,* 70(11), pp. 11184-11192.

Turner, W. J. N., Staino, A. & Basu, B., 2017. Residential HVAC fault detection using a system identification approach. *Energy and Buildings,* Volume 151, pp. 1-17.

Ugwiri, M. A. et al., 2022. Parameter estimation for fault detection and classification in centrifugal pumps. *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* pp. 1-6.

Unbehauen, H. & Rao, G. P., 1990. Continuous-time approaches to system identification—A survey. *Automatica,* 26(1), pp. 23-35.

Venkatasubramanian, V., Rengaswamy, R. & Kavuri, S. N., 2003b. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & chemical engineering,* 27(3), pp. 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N. & Yin, K., 2003c. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering,* 27(3), pp. 327-346.

Venkatasubramanian, V., Rengaswamy, R., Yin, K. & Kavuri, S. N., 2003a. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & chemical engineering,* 27(3), pp. 293-311.

Vileiniskis, M., Remenyte-Prescott, R. & Rama, D., 2016. A fault detection method for railway point systems. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 230(3), pp. 852-865.

Wang, G. et al., 2017. A Bayesian network model for prediction of weather-related failures in railway turnout systems. *Expert Systems with Applications,* Volume 69, pp. 247-256.

Wang, M. & Li, X., 2024. Application of artificial intelligence techniques in meat processing: A review. *Journal of Food Process Engineering,* 47(3), pp. 1-11.

Wang, N. et al., 2023. A self-adaptive phase-segmentation and health assessment framework for point machines. *IET Intelligent Transport Systems,* 17(4), pp. 730-743.

Wijewardhana, W. M. T. G., Senevirathne, E. H. & Abeykoon, A. M. H. S., 2020. Iterative Approach for Parameter Estimation in DC Motor Based Motion Systems. *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE),* pp. 134-141.

Wilhelm, Y., Reimann, P., Gauchel, W. & Mitschang, B., 2021. Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models. *Procedia Cirp,* Volume 99, pp. 278-283.

Williams, K., 2019. *Williams Rail Review - The role of the railway in Great Britain - Evidence paper.* [Online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955352/role-of-railway-evidence-paper-rail-review-document.pdf [Accessed 27 June 2023].

Yang, W. D., Zhang, J. L. & Gu, G., 2019. Parameter Estimation and Fault Diagnosis for Compensation Capacitators in ZPW-2000 Jointless Track Circuit. *2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS),* pp. 675-681.

Ye, J. et al., 2022. A vision-based method for line-side switch rail condition monitoring and inspection. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 236(8), pp. 986-996.

Young, P., 1974. Recursive Approaches to Time Series Analysis. *Bull. Inst. Math. Appl.,* Volume 10, pp. 209-224.

Young, P., 1981. Parameter estimation for continuous-time models—A survey. *Automatica,* 17(1), pp. 23-39.

Young, P., 1985. The Instrumental Variable Method: A Practical Approach to Identification and System Parameter Estimation. *IFAC Proceedings Volumes,* 18(5), pp. 1-15.

Young, P. C., 2012. *Recursive estimation and time-series analysis: an introduction.* 2nd ed. s.l.:Springer Science & Business Media.

Young, P. C., 2015. Refined instrumental variable estimation: Maximum likelihood optimization of a unified Box–Jenkins model. *Automatica,* Volume 52, pp. 35-46.

Zadeh, L. A., 1965. Fuzzy sets. *Information and control,* 8(3), pp. 338-353.

Zadeh, L. A., 1975a. The concept of a linguistic variable and its application to approximate reasoning—I. *Information sciences,* 8(3), pp. 199-249.

Zadeh, L. A., 1975b. The concept of a linguistic variable and its application to approximate reasoning—II. *Information sciences,* 8(4), pp. 301-357.

Zadeh, L. A., 1975c. The concept of a linguistic variable and its application to approximate reasoning-III. *Information Sciences,* 9(1), pp. 43-80.

Zattoni, E., 2006. Detection of incipient failures by using an H2-norm criterion: Application to railway switching points. *Control Engineering Practice,* 14(8), pp. 885-895.

Zhang, P. et al., 2018. Fault Diagnosis of High-Speed Railway Turnout Based on Convolutional Neural Network. *2018 24th International Conference on Automation and Computing (ICAC),* pp. 1-6.

Zhang, Y. & Jiang, J., 2008. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control,* 32(2), pp. 229-252.

Zhang, Z., Xu, B., Ma, L. & Geng, S., 2014. Parameter estimation of a railway vehicle running bogie using extended Kalman filter. *Proceedings of the 33rd Chinese Control Conference,* pp. 3393-3398.

Zhao, H. et al., 2018. Fault diagnosis for rail turnout system based on improved introspective CBR method. *2018 13th World Congress on Intelligent Control and Automation (WCICA),* pp. 989-994.

Zhao, L. H., Zhang, C. L., Qiu, K. M. & Li, Q. L., 2013. *A fault diagnosis method for the tuning area of jointless track circuits based on a neural network.* s.l., Institution of Mechanical Engineers, pp. 333-343.

Zhao, N. et al., 2017. *Fault Analysis of the Switch and Point Machine to Improve Railway Reliability.* London, s.n.

Zhiwei, G., Cecati, C. & Ding, S. X., 2015b. A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE transactions on industrial electronics,* 62(6), pp. 3768-3774.

Zhou, F. et al., 2002a. *Remote condition monitoring and validation of railway points.,* pp. 221-230.

Zhou, F. B. et al., 2002b. Remote condition monitoring for railway point machine. *ASME/IEEE Joint Railroad Conference,* pp. 103-108.

Zhou, F. et al., 2001. *Condition monitoring and validation of railway point machines.,* pp. 6/1-6/7.

Zhou, F. et al., 2016. Fault diagnosis of high-speed railway turnout based on support vector machine. *2016 IEEE International Conference on Industrial Technology (ICIT),* pp. 1539-1544.

Zoljic-Beglerovic, S., Golkani, M. A., Steinberger, M. & Horn, M., 2018. Robust Parameter Identification for Railway Suspension Systems. *2018 15th International Workshop on Variable Structure Systems (VSS),* pp. 432-437.

APPENDIX 1 TERMINOLOGIES PREVIOUSLY USED TO DESCRIBE COMMON FAULT MODES

Common fault modes (corresponding with Table 2-1)	Terminologies found in the selected publications
Obstruction	Changing the fulcrum point; Ballast/ice obstruction; Switch machine idle; Fail to lock; Fail to start; Fail to switch; Mechanical jam; Jamming of the throw bar; Increasing obstruction size; Getting stuck due to snow; Lock failure; Turnout blocked; Ice block; Slide chair obstructed.
Improper Sliding Surface	Resistance increase due to temperature; Switch moving parts lacking oil/lubrication; Switch operation has resistance; Inadequate lubrication; Changes in the friction elements; Excessive friction (from a hydraulic device); Improper position of the slide chairs; Dry/contamination/deterioration/frozen/defective slide chairs; Loss/lacking lubrication; Missing rollers; Abnormal fluctuation; Conversion current exceeded; Misalignment of slide chairs; Large moving resistance.
Detection-related	The indicating current is missing; Switch cannot detect; Indicating rod block in the gap; Open-phase protection device fault; Indicating circuit diode damage; Lack/out of adjustment of detector bars; Indication circuit cannot work; Detection contact failure; Microswitch defective; Abnormal impedance in the indicating circuit; Abnormal open-phase detection device; Slowly switching of detection contacts; Overlong release time.
Overdrive	Overdriving; Tight lock; Backdrive overdriving; Locking mechanism tightened; Switch rail rebounds while locking (with noise and hunting); Turnout unlocking; Hard to release; Inadequate/lack/out of adjustment of lock/drive bars; Adjuster bolts misalignment; Additional mechanical resistance in lock phase; Facing point lock out of adjustment.
P-way Problems	Misalignment of stock rails; Movement of the sleepers and rails; Switch operation has resistance; Changes in weather/stiffness in switch rails; Curvature of a tongue rail; Switch point creepage; Broken stretcher bar; Loose lock-pin; Incorrect installation; Run-through damage by vehicles; Track maintenance; Ballast voiding; Rails are not in proper positions.

Common fault modes (corresponding with Table 2-1)	Terminologies found in the selected publications
Faults in Drive Circuit	Increase in winding resistance; Changes in rotor blade mass; Motor problems; Abnormal impedance; Bad/poor contact in the drive circuit; Crushed cable; Low tension on motor brush; Short-circuit in motor winding; Machine idling; Reduced input voltage.
Underdrive	Lack/out of adjustment of lock/drive bars; Locking mechanism loosened; Clearance between rails; Slackened nut; Switch machine idle; Inadequate adjustment; Fail to lock; Backdrive slackened off; Adjuster bolts misalignment; Facing point lock out of adjustment.
Electrical Failures	Major electrical system failure; Relay in start circuit fails; Supply interruption; Breakdown of electronic components; Open-phase protection device fails; Supply open phase; Electric relay 2DJ turnout failure; Open start-up circuit; Release failure; Drive circuit defective (from fuse, relay, contact, motor brushes).
Mechanical Faults in Point Machine	Point machine notches stuck; Damaged bearings; Unstable work of the friction clutch; Slit up of point machine's bearings; Friction change; Gear misalignments; Intermittent motion from clutch slipping.

APPENDIX 2 PARAMETER ESTIMATION RESULTS OF THE HW55S TURNOUT SYSTEM

This section presents all of the parameter estimation results from both N-to-R and R-to-N training operations of the HW55S turnout. The intention is to demonstrate that the relationships between qualitative parameter changes and the associated fault modes are identical for both N-to-R and R-to-N operations. As a result, this enabled the same knowledge-base parameter change table (i.e. Table 5-5) to be applicable to generate diagnostic fuzzy rules for the R-to-N HW55S operations as well.

A2.1 The estimation results of N-to-R HW55S operations

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated overdrive fault conditions are shown in Figure A2-1 and their estimation results are shown in Figure A2-2.

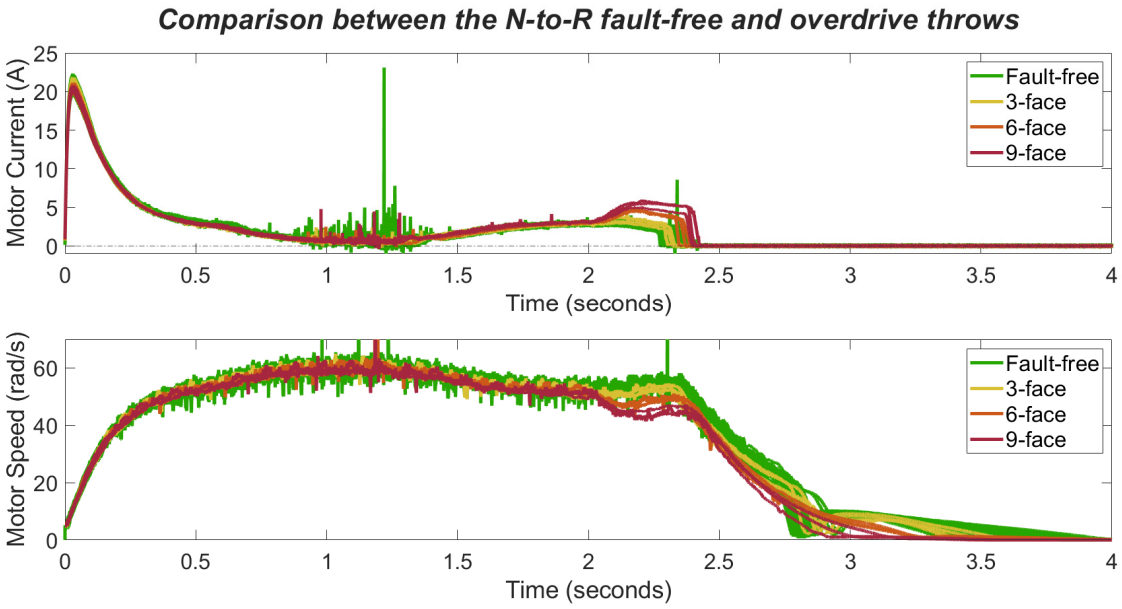


Figure A2-1 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated overdrive fault conditions

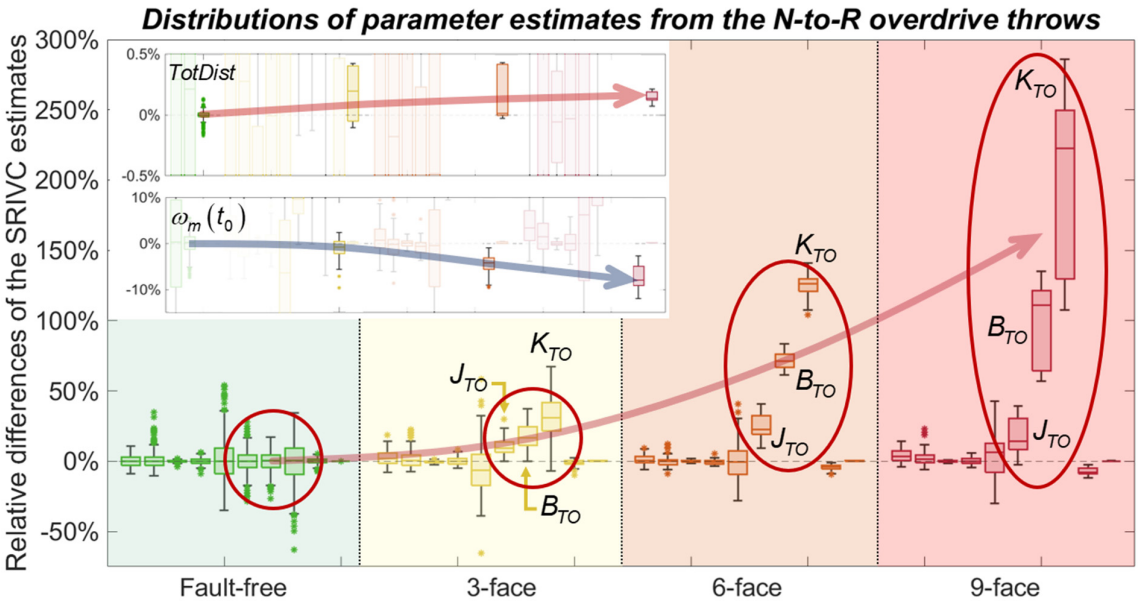


Figure A2-2 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated overdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout is operated in N-to-R direction with the simulated unlock difficulty fault conditions are shown in Figure A2-3 and their estimation results are shown in Figure A2-4.

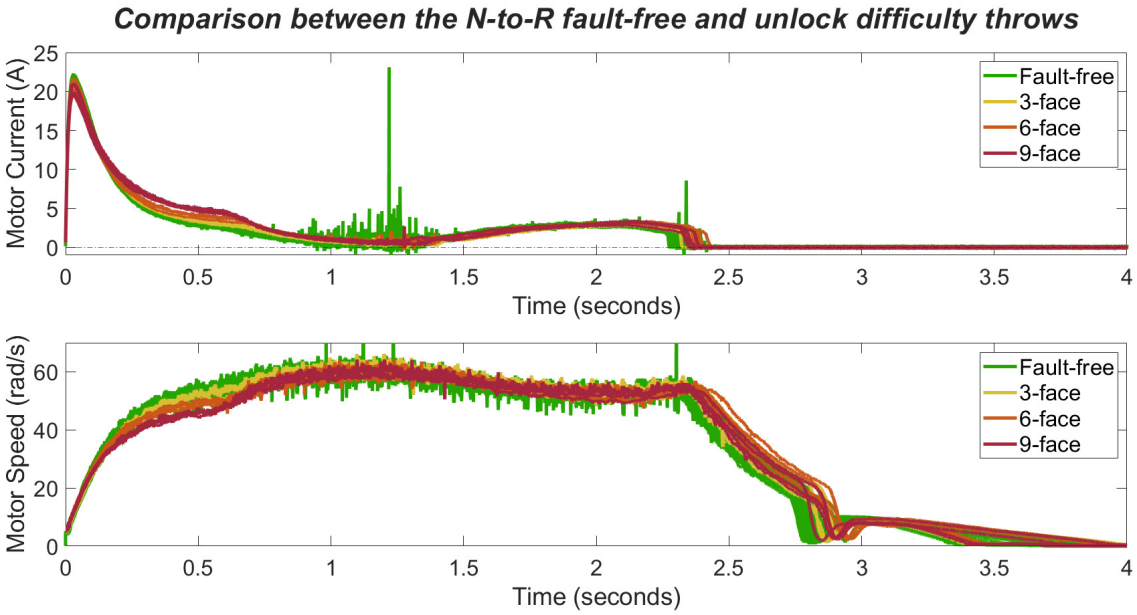


Figure A2-3 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated unlock difficulty fault conditions

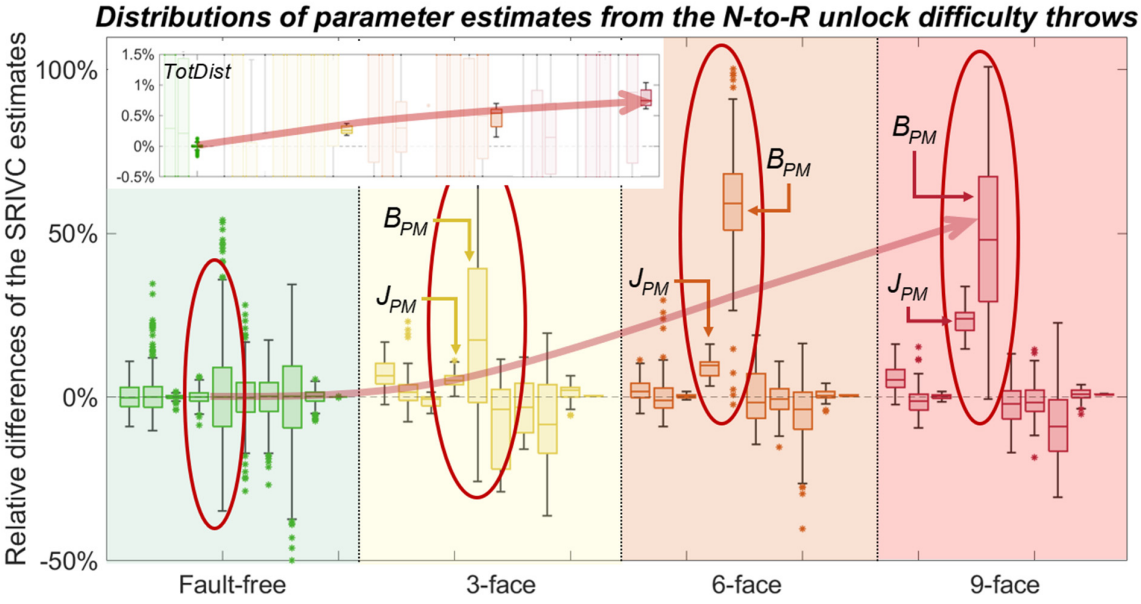


Figure A2-4 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated unlock difficulty fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated obstruction fault conditions are shown in Figure A2-5 and their estimation results are shown in Figure A2-6.

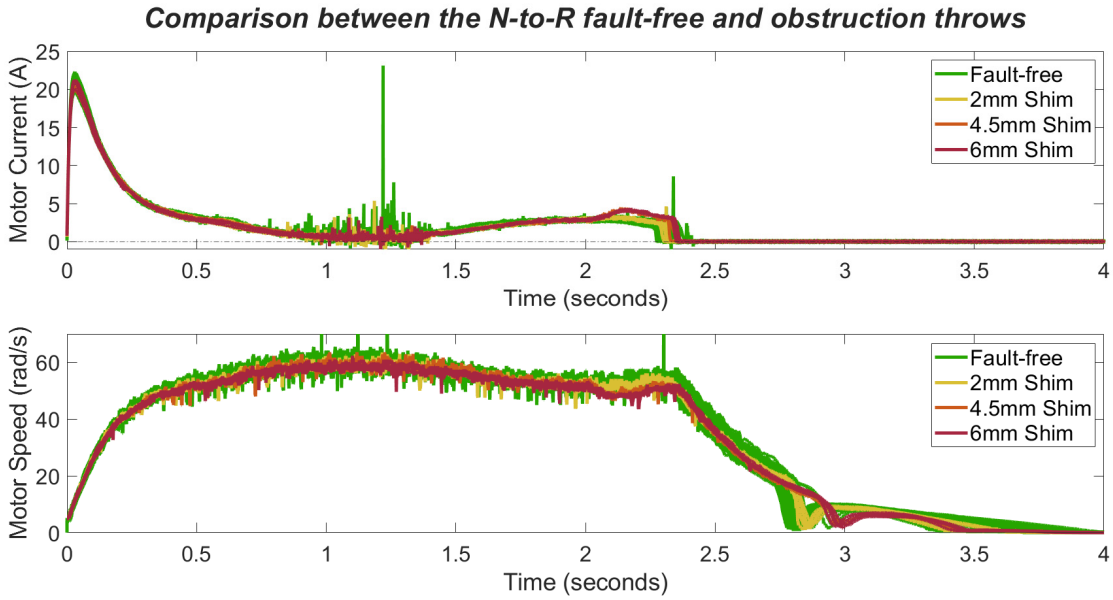


Figure A2-5 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated obstruction fault conditions

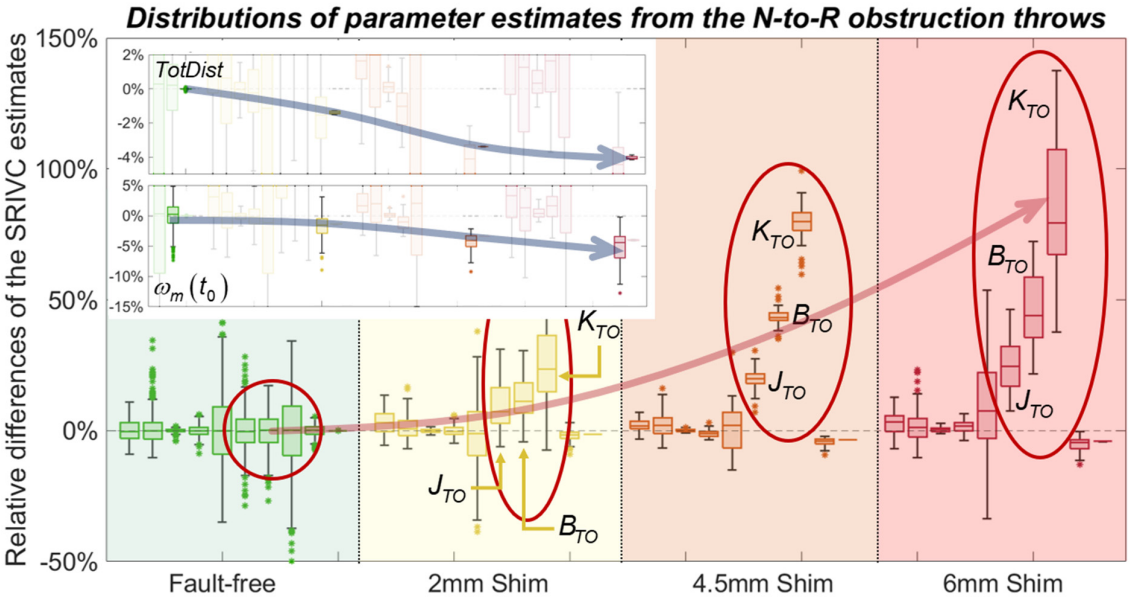


Figure A2-6 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated obstruction fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated OPP-obstruction faults are shown in Figure A2-7 and their estimation results are shown in Figure A2-8.

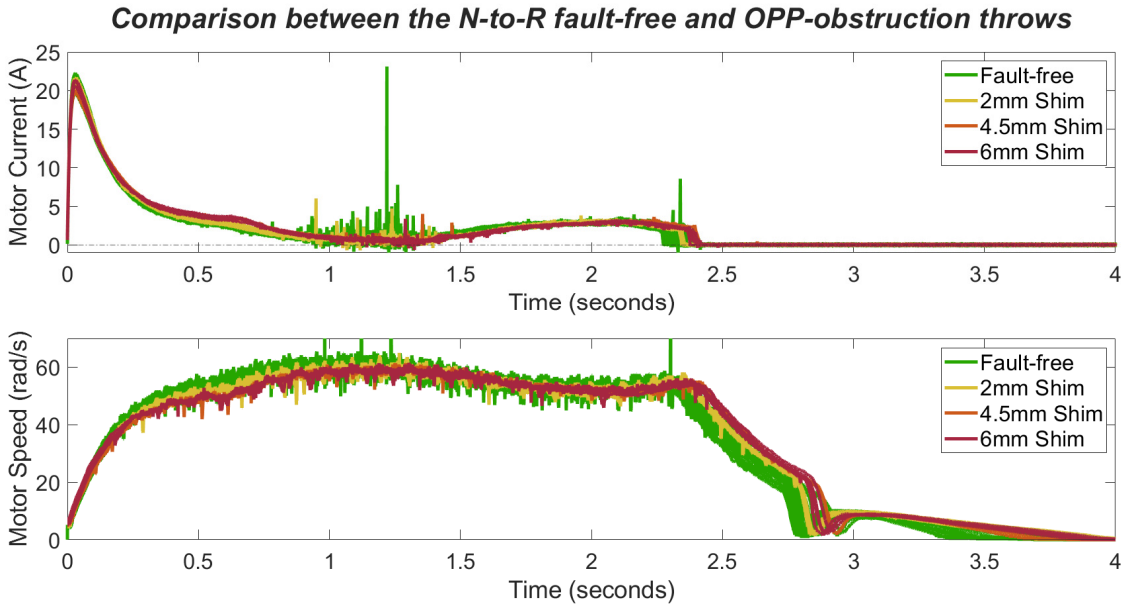


Figure A2-7 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated OPP-obstruction fault conditions

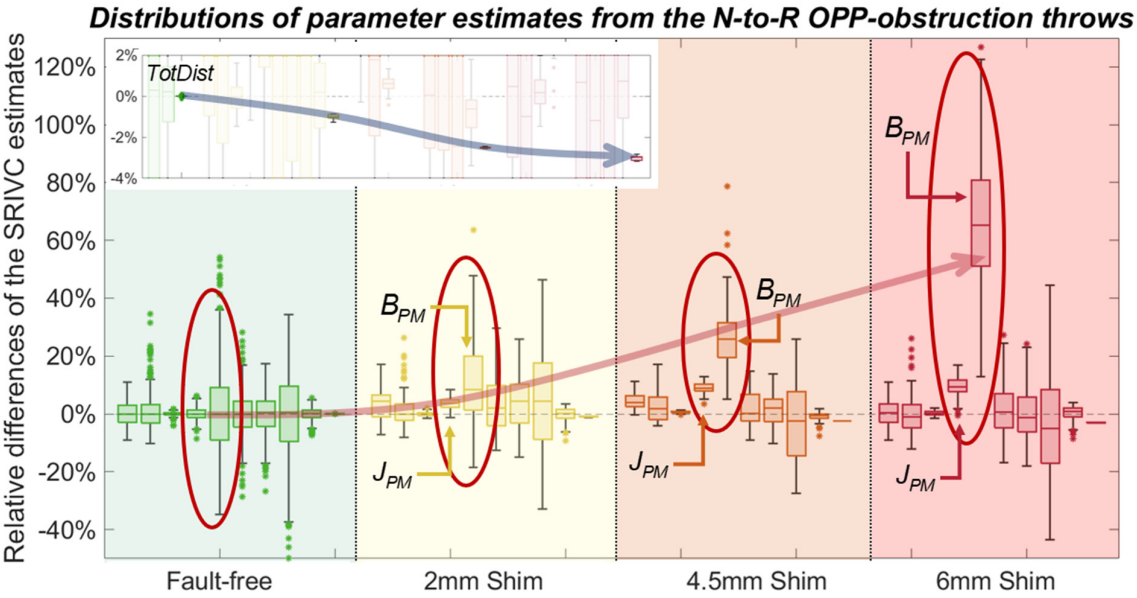


Figure A2-8 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated OPP-obstruction fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated underdrive fault conditions are shown in Figure A2-9 and their estimation results are shown in Figure A2-10.

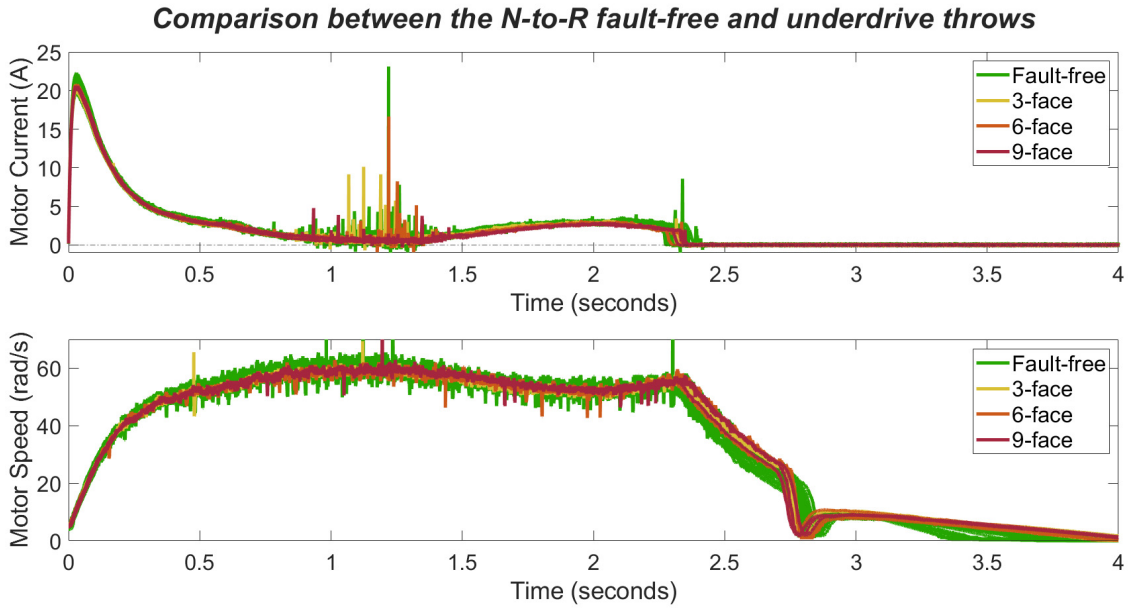


Figure A2-9 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated underdrive fault conditions

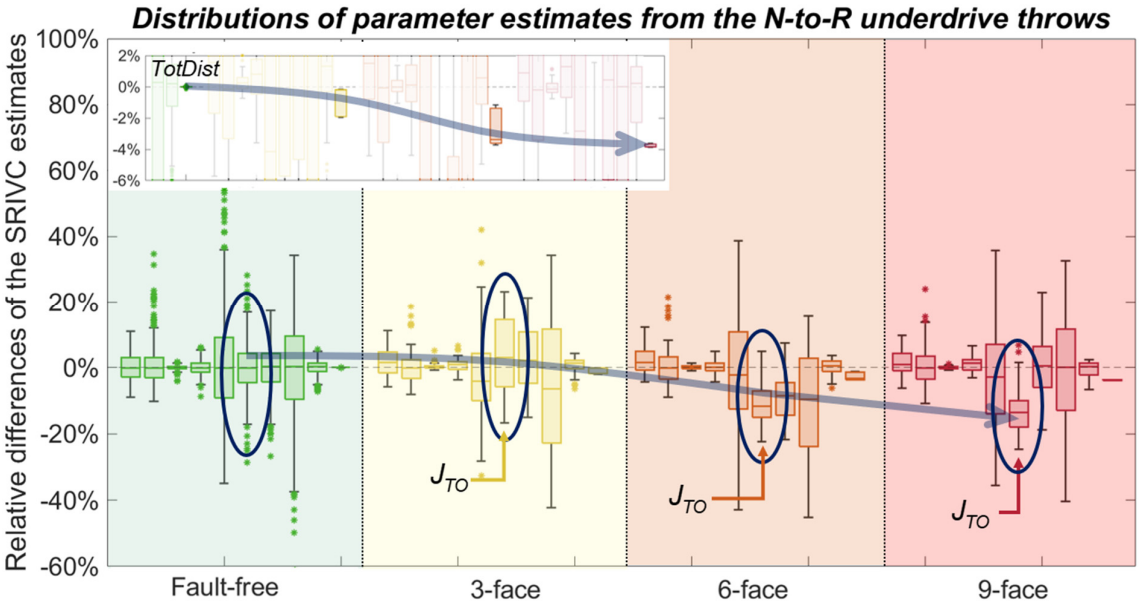


Figure A2-10 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated underdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated OPP-underdrive faults are shown in Figure A2-11 and their estimation results are shown in Figure A2-12.

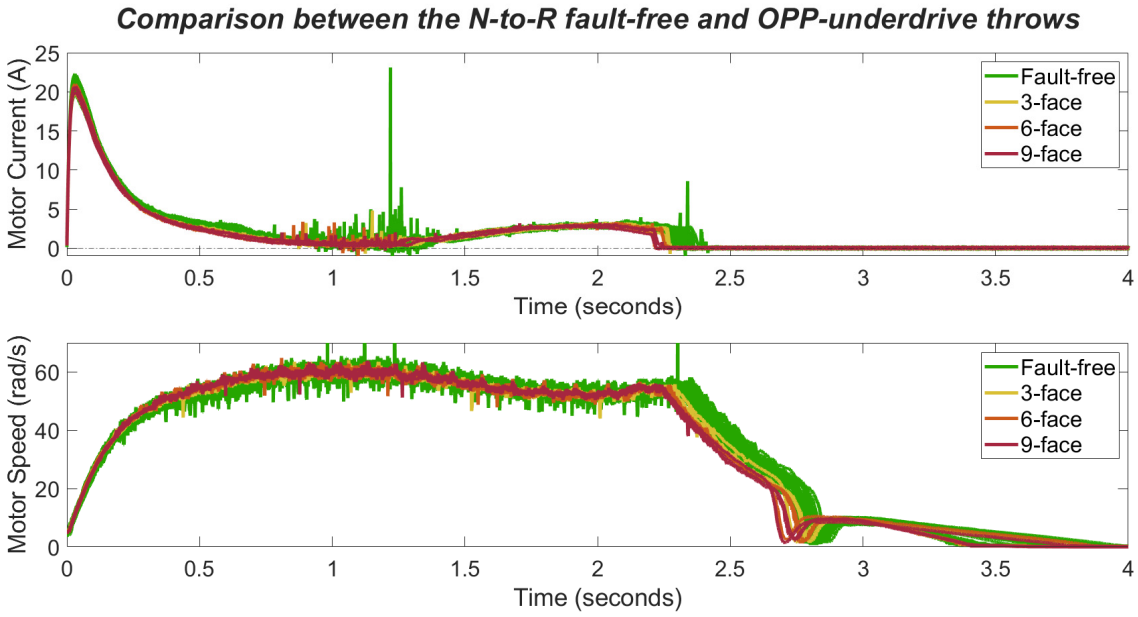


Figure A2-11 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated OPP-underdrive fault conditions

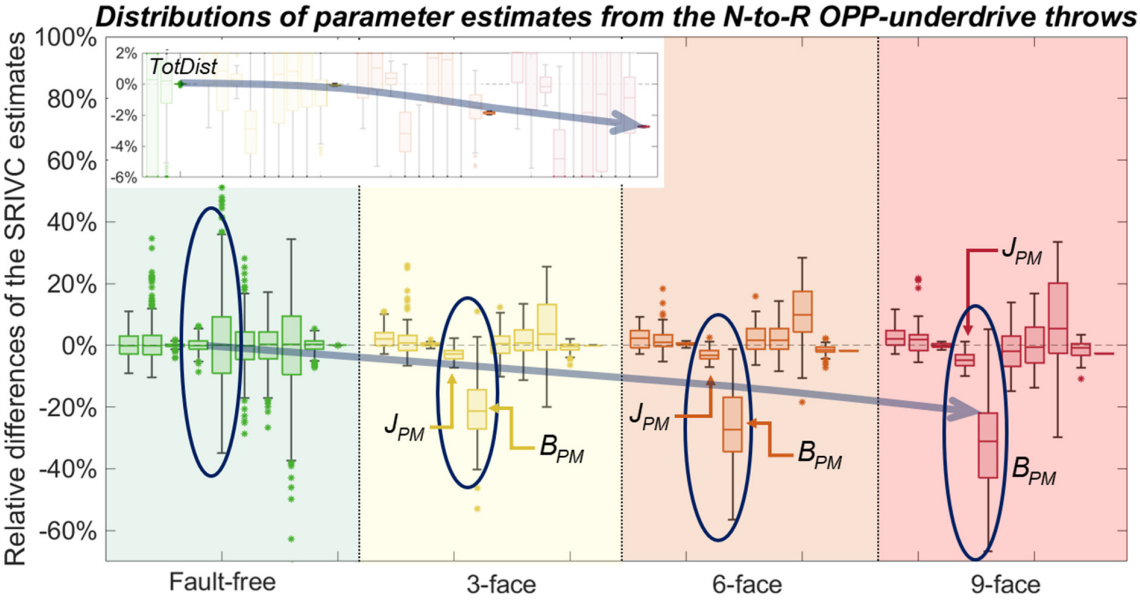


Figure A2-12 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated OPP-underdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated abnormal input voltage faults are shown in Figure A2-13 and their estimation results are shown in Figure A2-14.

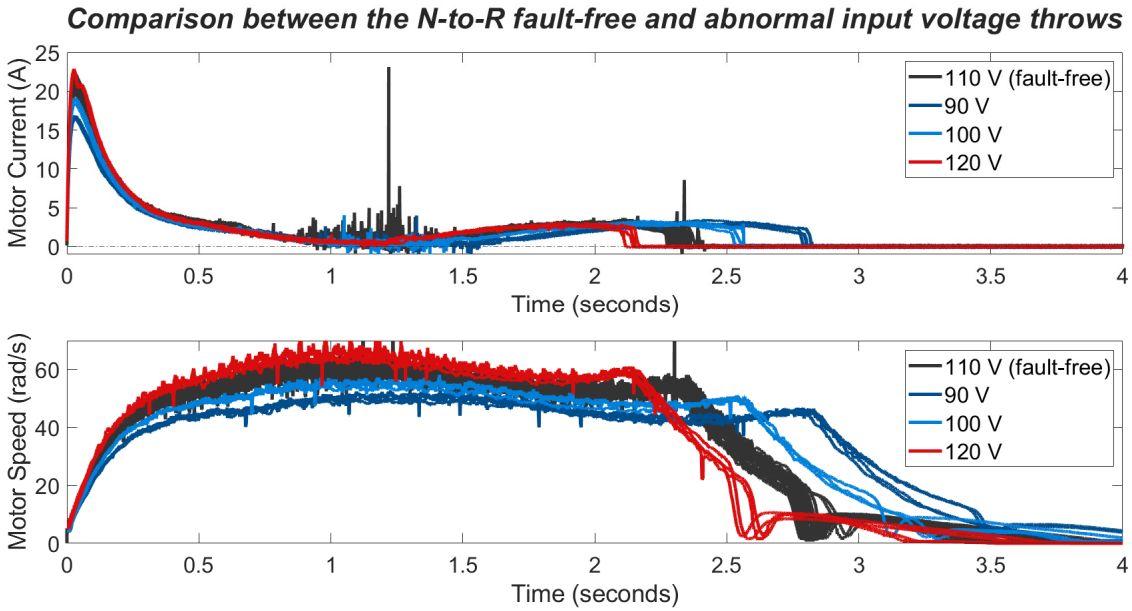


Figure A2-13 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated abnormal voltage fault conditions

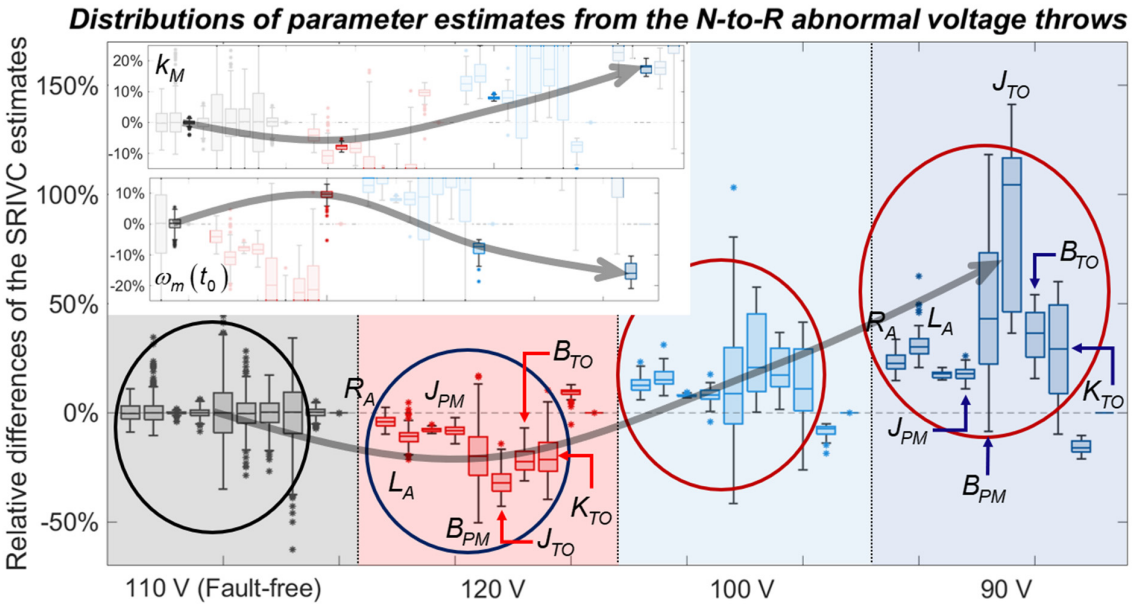


Figure A2-14 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated abnormal voltage fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in N-to-R direction with the simulated improper sliding surface faults are shown in Figure A2-15 and their estimation results are shown in Figure A2-16.

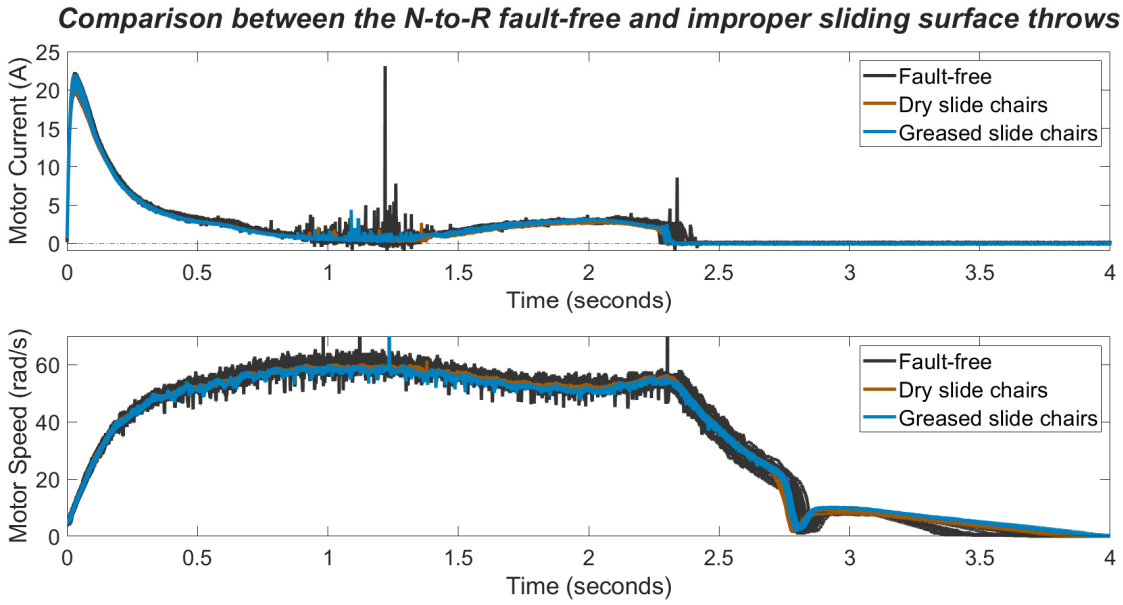


Figure A2-15 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in N-to-R direction with simulated improper surface fault conditions

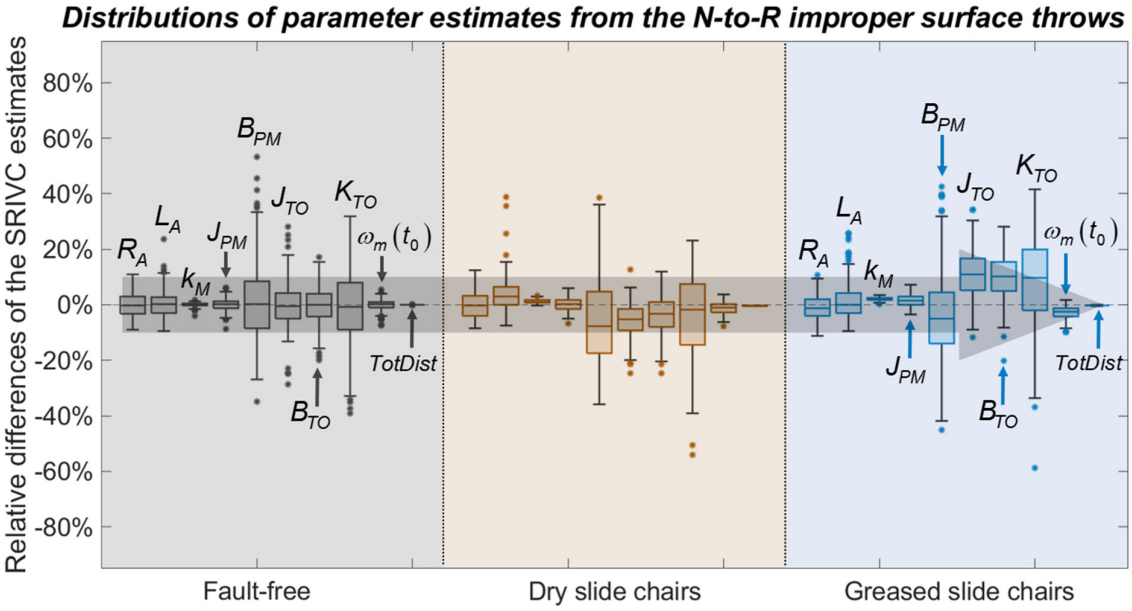


Figure A2-16 Boxplot distributions of the estimation results when the HW55S turnout system is operated in N-to-R direction with simulated improper surface fault conditions

A2.2 The estimation results of R-to-N HW55S operations

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated overdrive fault conditions are shown in Figure A2-17 and their estimation results are shown in Figure A2-18.

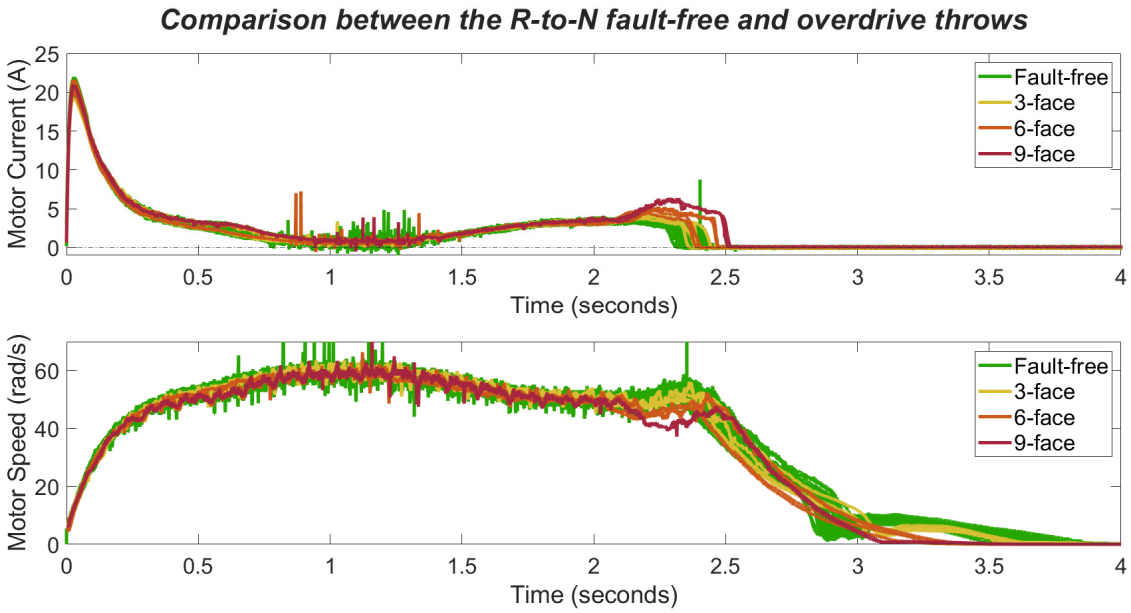


Figure A2-17 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated overdrive fault conditions

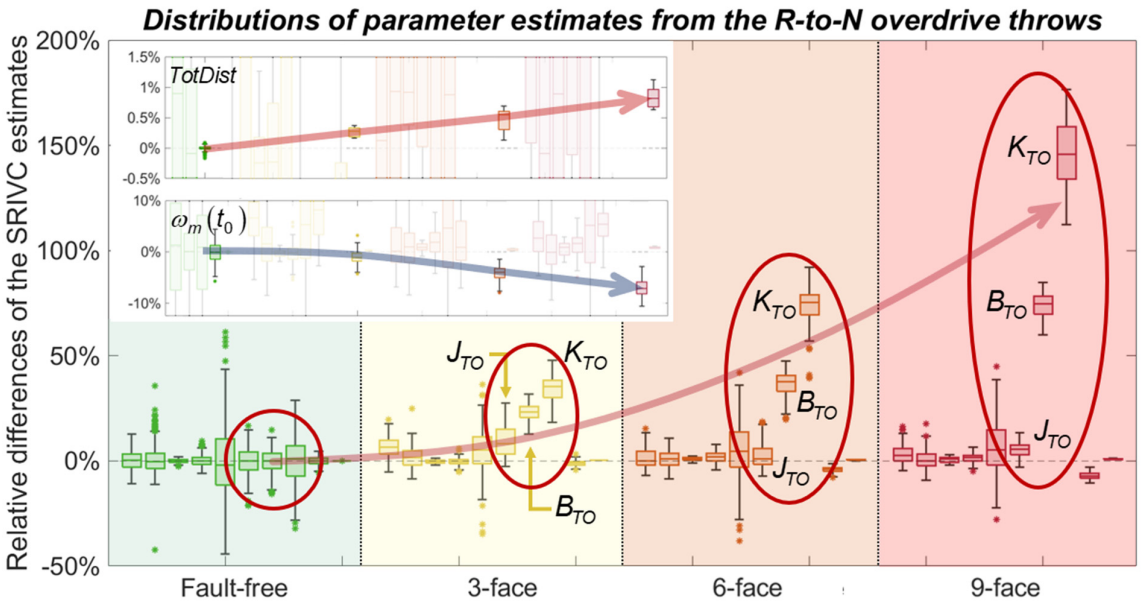


Figure A2-18 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated overdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout is operated in R-to-N direction with the simulated unlock difficulty fault conditions are shown in Figure A2-19 and their estimation results are shown in Figure A2-20.

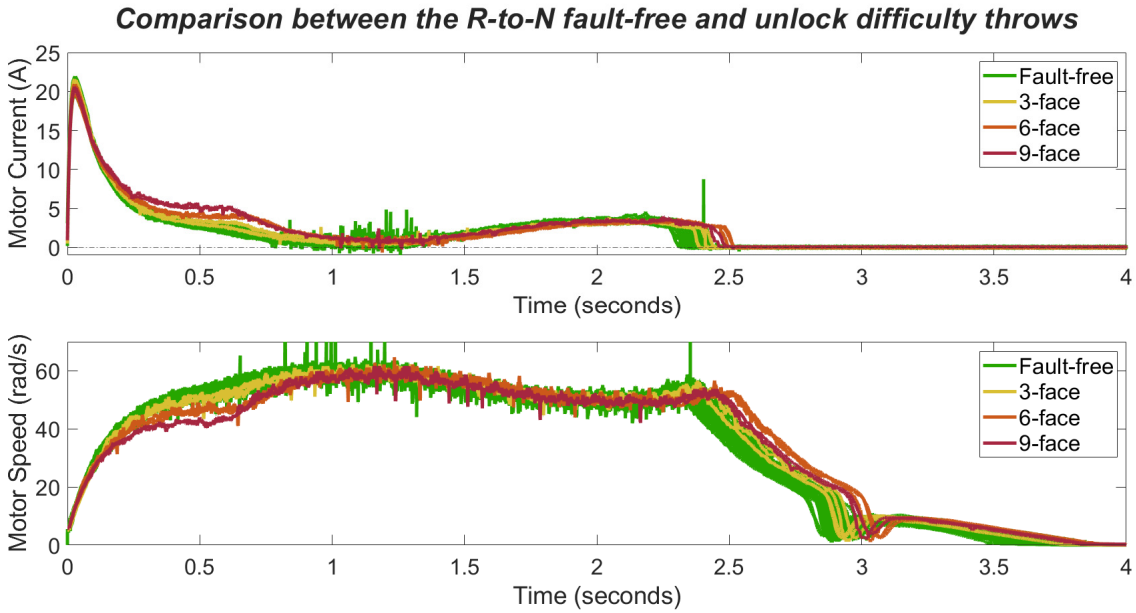


Figure A2-19 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated unlock difficulty fault conditions

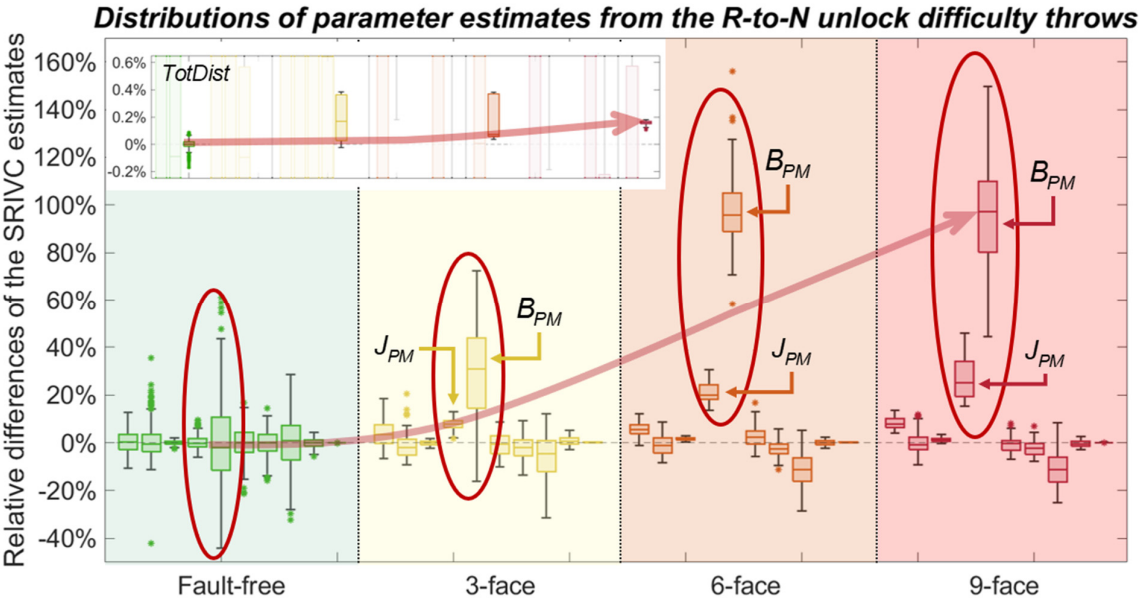


Figure A2-20 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated unlock difficulty fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated obstruction fault conditions are shown in Figure A2-21 and their estimation results are shown in Figure A2-22.

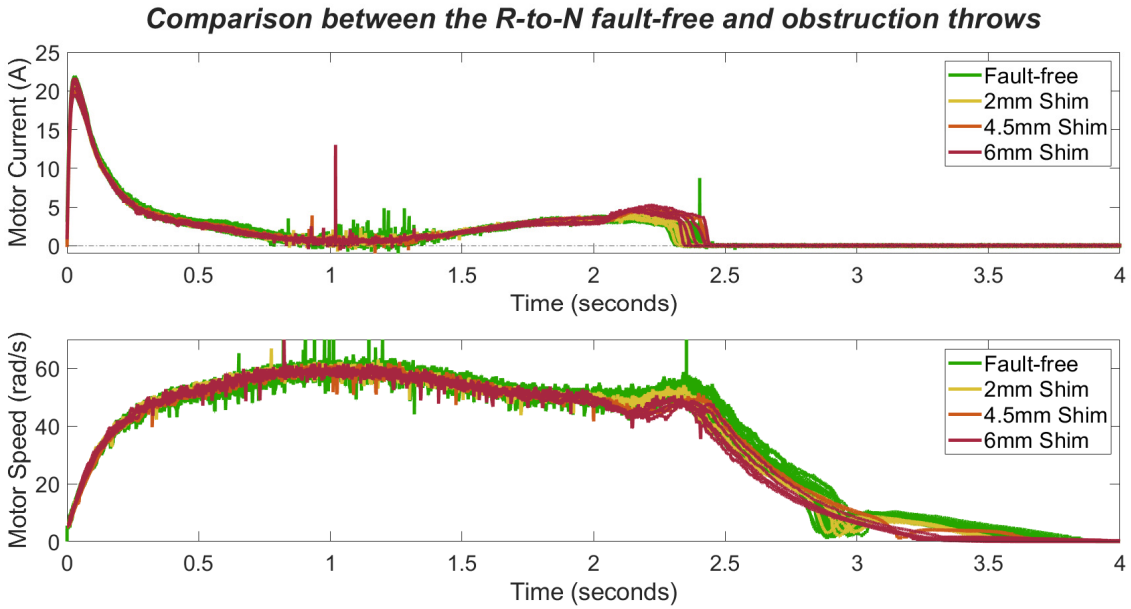


Figure A2-21 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated obstruction fault conditions

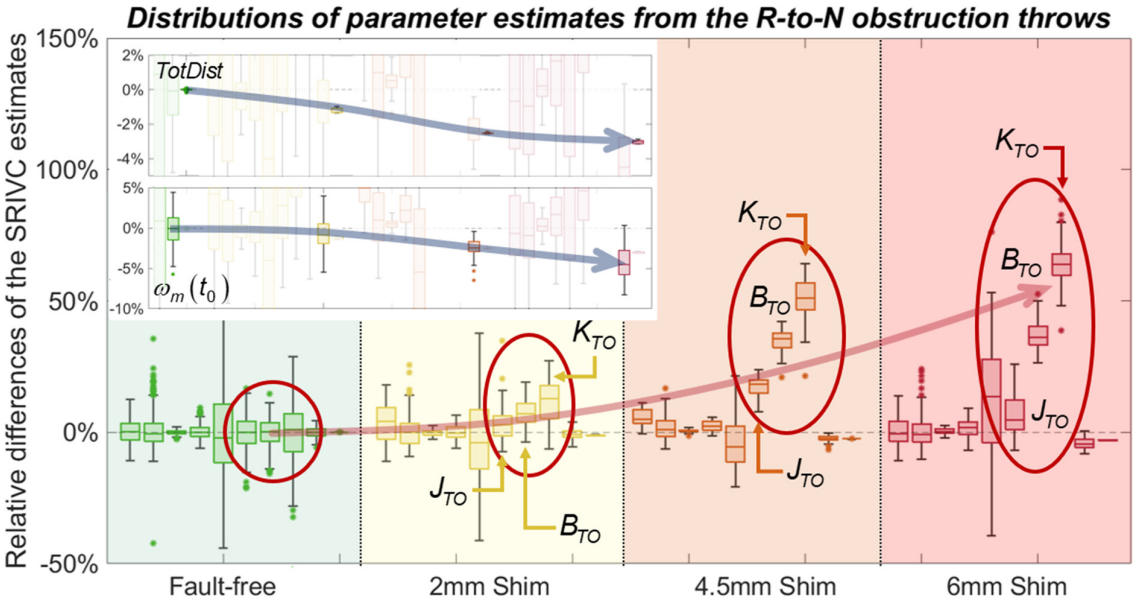


Figure A2-22 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated obstruction fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated OPP-obstruction faults are shown in Figure A2-23 and their estimation results are shown in Figure A2-24.

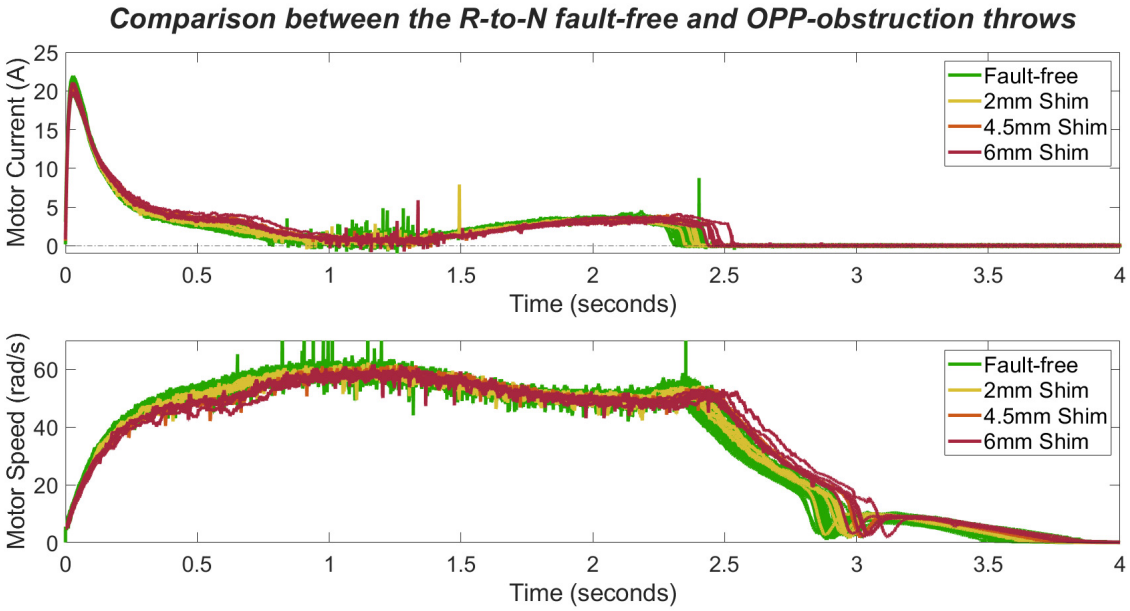


Figure A2-23 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated OPP-obstruction fault conditions

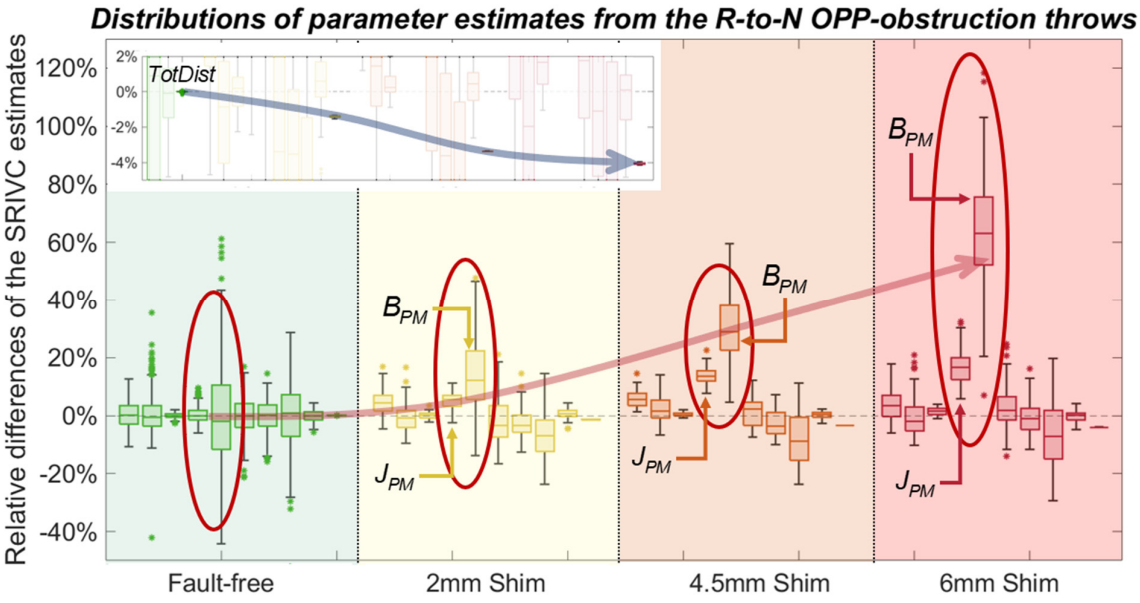


Figure A2-24 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated OPP-obstruction fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated underdrive fault conditions are shown in Figure A2-25 and their estimation results are shown in Figure A2-26.

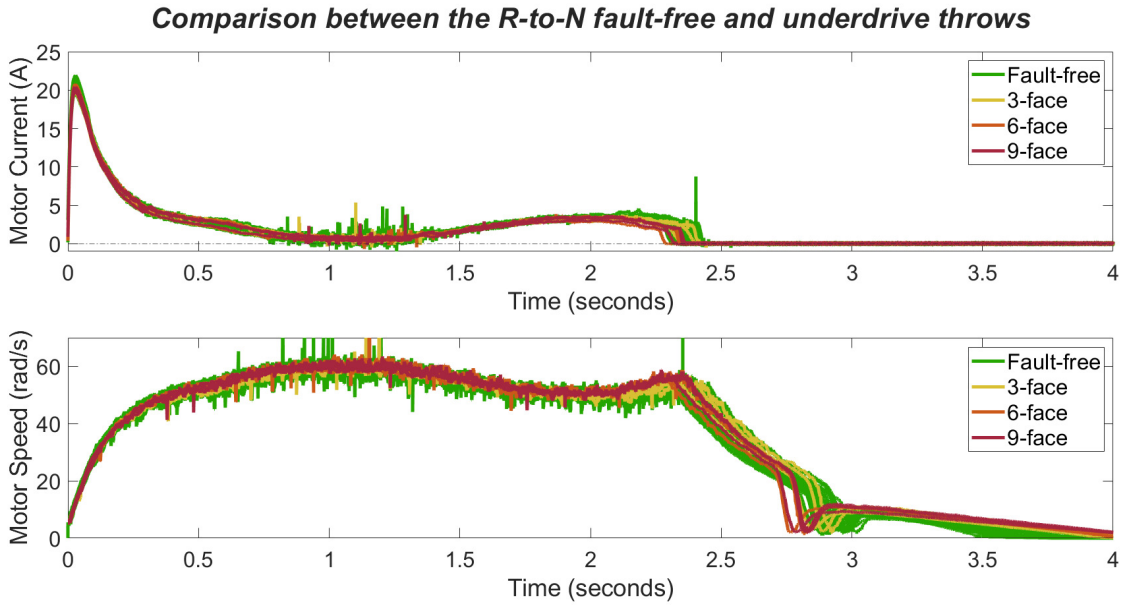


Figure A2-25 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated underdrive fault conditions

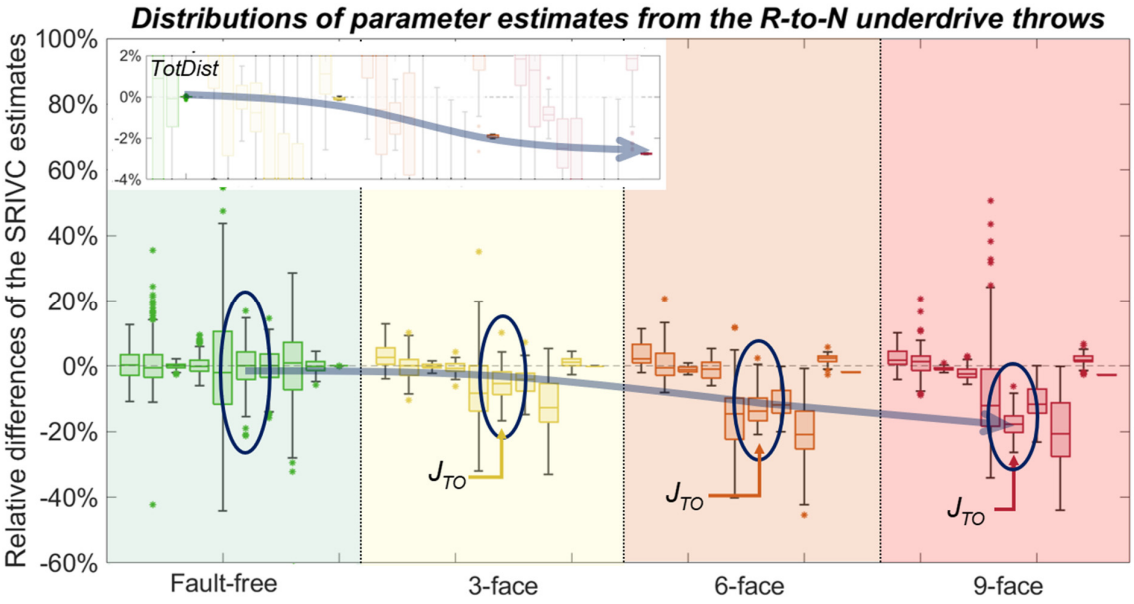


Figure A2-26 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated underdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated OPP-underdrive faults are shown in Figure A2-27 and their estimation results are shown in Figure A2-28.

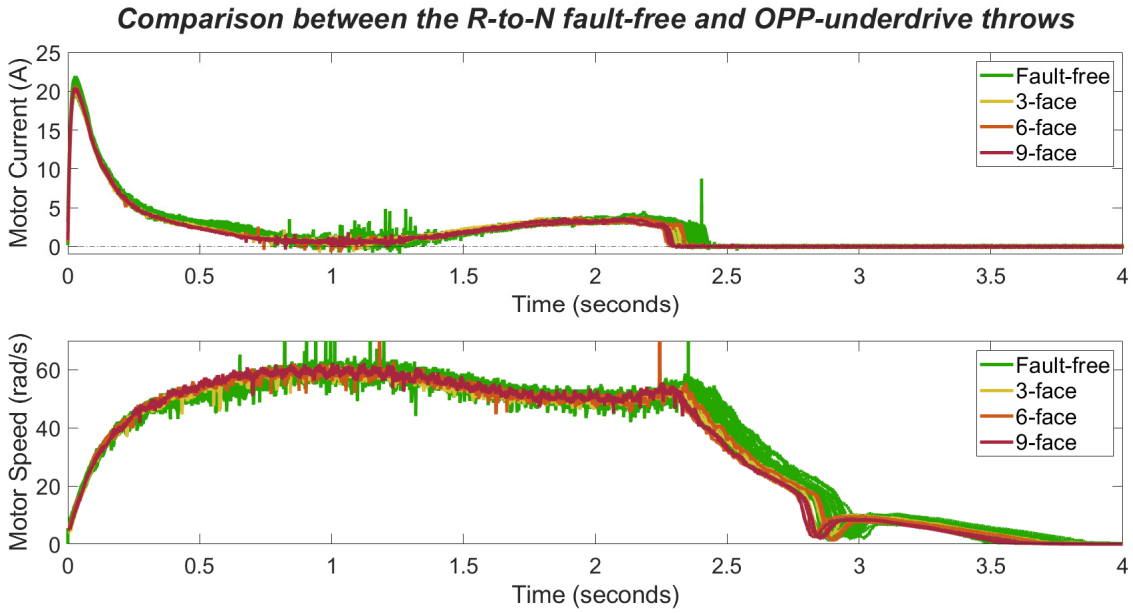


Figure A2-27 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated OPP-underdrive fault conditions

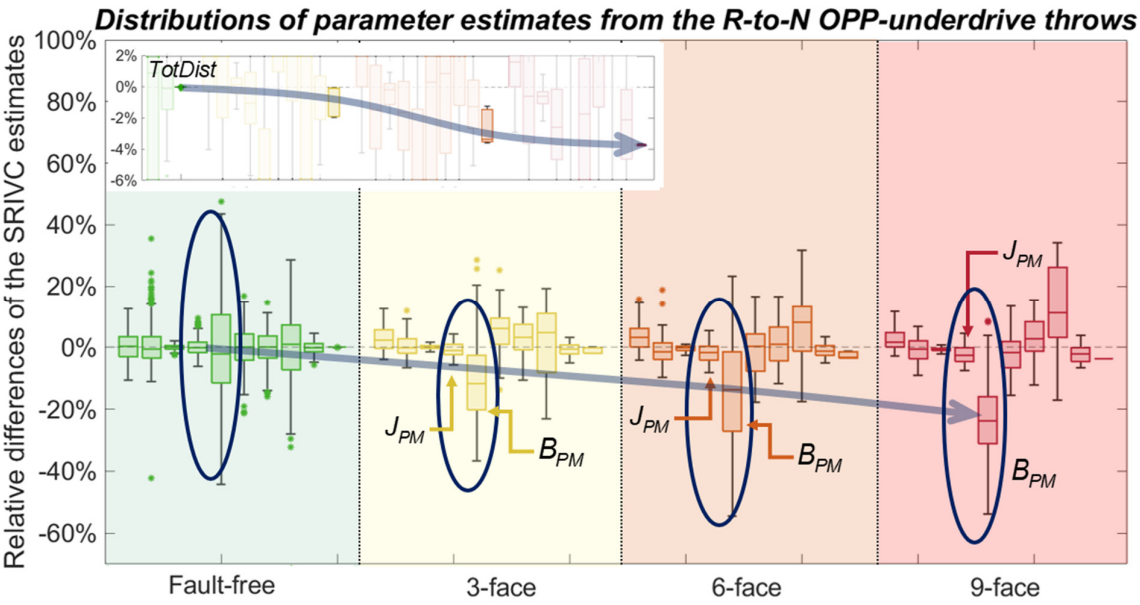


Figure A2-28 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated OPP-underdrive fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated abnormal input voltage faults are shown in Figure A2-29 and their estimation results are shown in Figure A2-30.

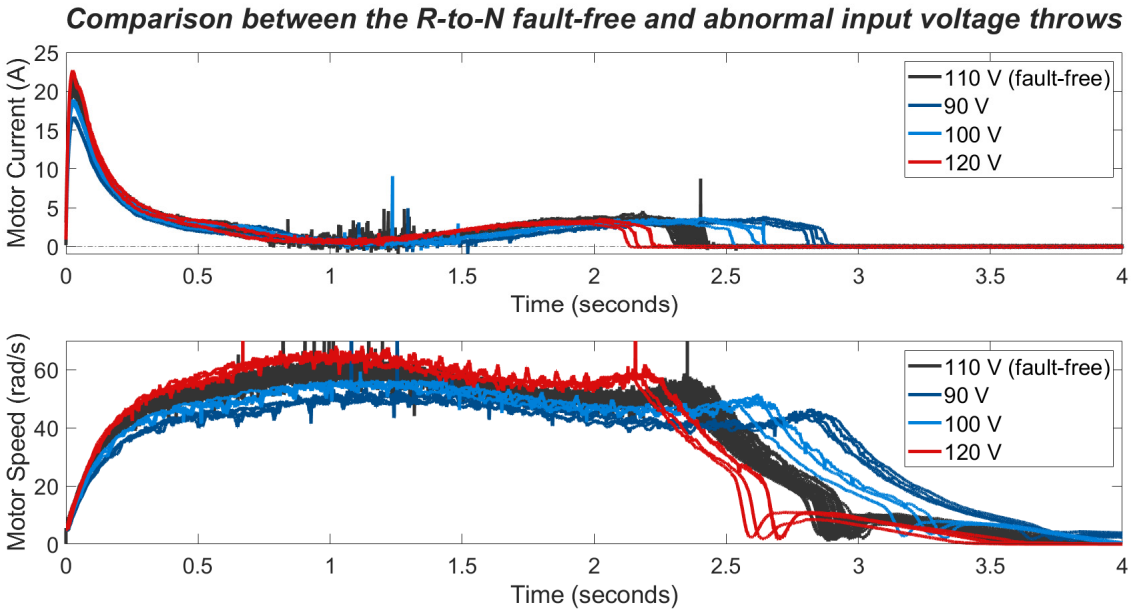


Figure A2-29 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated abnormal voltage fault conditions

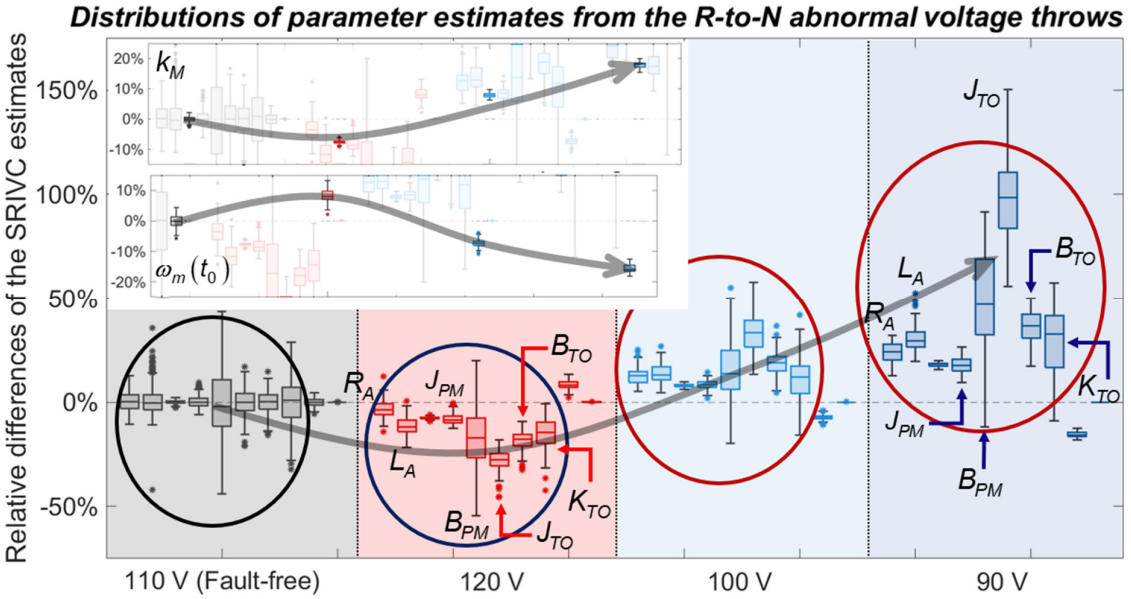


Figure A2-30 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated abnormal voltage fault conditions

The measurements of the motor current and speed when the HW55S turnout system is operated in R-to-N direction with the simulated improper sliding surface faults are shown in Figure A2-31 and their estimation results are shown in Figure A2-32.

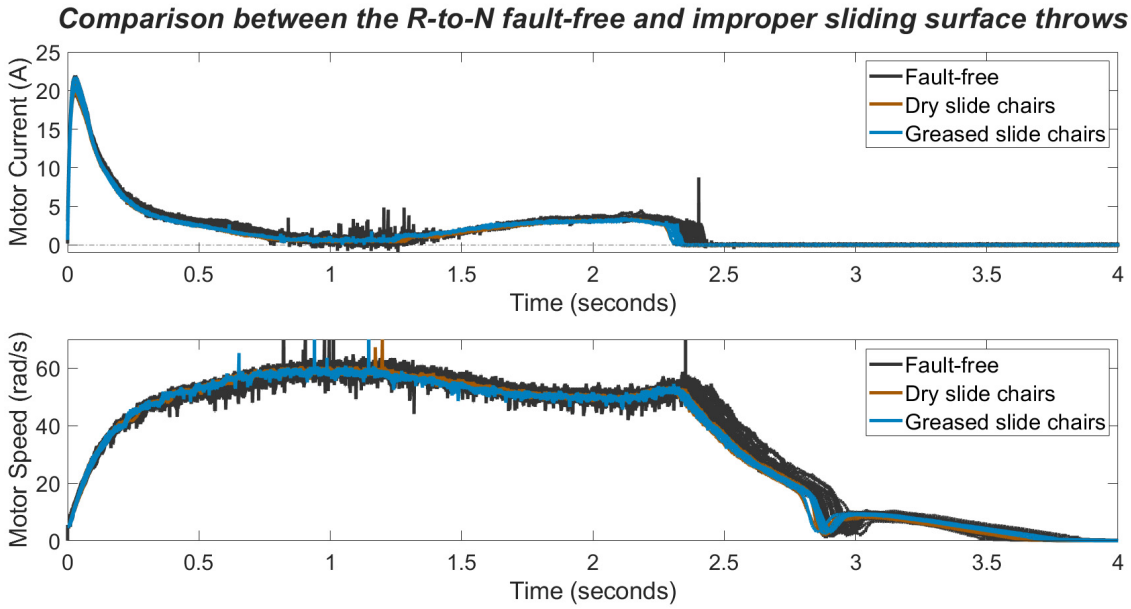


Figure A2-31 Measurements of motor current and motor speed obtained from the HW55S turnout system operated in R-to-N direction with simulated improper surface fault conditions

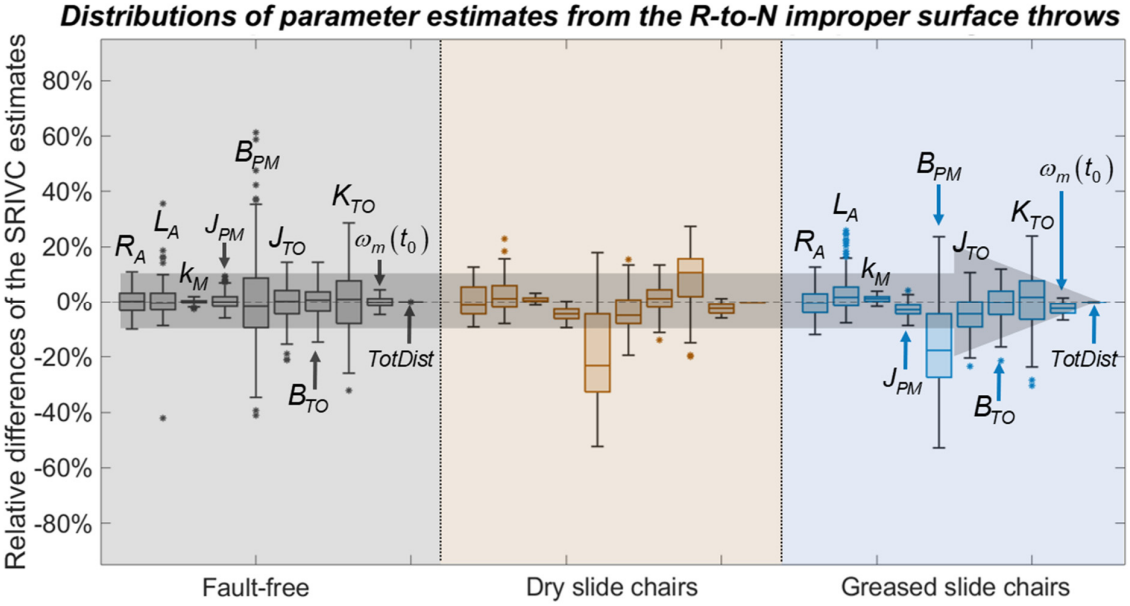


Figure A2-32 Boxplot distributions of the estimation results when the HW55S turnout system is operated in R-to-N direction with simulated improper surface fault conditions

APPENDIX 3 FUZZY FAULT DIAGNOSIS RULES

Based on the fuzzy rule generation principle exemplified in Section 6.2.4, all of the eight fuzzy rules are provided in this appendix.

The fuzzy rule used to identify overdrive faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Overdrive						↑	↑↑	↑↑	↓	↑↑

Fuzzy rule to detect the occurrence of "overdrive" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	ANY
AND	B_{PM}	is	ANY	AND	J_{TO}	is	NOT Low
AND	B_{TO}	is	High	AND	K_{TO}	is	High
AND	$\omega_m(t_0)$	is	NOT High	AND	*TotDist*	is	High
Then	State	is	Overdrive				

The fuzzy rule used to identify unlock difficulty faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Unlock Difficulty				↑↑	↑					↑↑

Fuzzy rule to detect the occurrence of "unlock difficulty" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	High
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	ANY
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	High
Then	State	is	UnlockDifficulty				

The fuzzy rule used to identify obstruction faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Obstruction						↑	↑↑	↑↑	↓	↓↓

Fuzzy rule to detect the occurrence of "obstruction" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	ANY
AND	B_{PM}	is	ANY	AND	J_{TO}	is	NOT Low
AND	B_{TO}	is	High	AND	K_{TO}	is	High
AND	$\omega_m(t_0)$	is	NOT High	AND	*TotDist*	is	Low
Then	State	is	Obstruction				

The fuzzy rule used to identify OPP-obstruction faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
OPP-obstruction				↑↑	↑					↓↓

Fuzzy rule to detect the occurrence of "OPP-obstruction" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	High
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	ANY
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	Low
Then	State	is	OPP-Obstruction				

The fuzzy rule used to identify underdrive faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Underdrive						↓				↓↓

Fuzzy rule to detect the occurrence of "underdrive" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	ANY
AND	B_{PM}	is	ANY	AND	J_{TO}	is	NOT High
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	Low
Then	State	is	Underdrive				

The fuzzy rule used to identify OPP-underdrive faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
OPP-underdrive				↓	↓					↓↓

Fuzzy rule to detect the occurrence of "OPP-underdrive" faults

If	R_A	is	ANY	AND	L_A	is	ANY
AND	k_M	is	ANY	AND	J_{PM}	is	NOT High
AND	B_{PM}	is	NOT High	AND	J_{TO}	is	ANY
AND	B_{TO}	is	ANY	AND	K_{TO}	is	ANY
AND	$\omega_m(t_0)$	is	ANY	AND	*TotDist*	is	Low
Then	State	is	OPP-Underdrive				

The fuzzy rule used to identify overvoltage faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Overvoltage	↓	↓	↓↓	↓↓	↓	↓	↓↓	↓	↑↑	↑

Fuzzy rule to detect the occurrence of "overvoltage" faults

If	R_A	is	NOT High	AND	L_A	is	NOT High
AND	k_M	is	Low	AND	J_{PM}	is	Low
AND	B_{PM}	is	NOT High	AND	J_{TO}	is	NOT High
AND	B_{TO}	is	Low	AND	K_{TO}	is	NOT High
AND	$\omega_m(t_0)$	is	High	AND	*TotDist*	is	NOT Low
Then	State	is	Overvoltage				

The fuzzy rule used to identify undervoltage faults for the HW55S turnout system

Fault mode	***Turnout parameters***									
	R_A	L_A	k_M	J_{PM}	B_{PM}	J_{TO}	B_{TO}	K_{TO}	$\omega_m(t_0)$	*TotDist*
Undervoltage	↑	↑	↑↑	↑↑	↑	↑	↑↑	↑	↓↓	↓

Fuzzy rule to detect the occurrence of "undervoltage" faults

If	R_A	is	NOT Low	AND	L_A	is	NOT Low
AND	k_M	is	High	AND	J_{PM}	is	High
AND	B_{PM}	is	NOT Low	AND	J_{TO}	is	NOT Low
AND	B_{TO}	is	High	AND	K_{TO}	is	NOT Low
AND	$\omega_m(t_0)$	is	Low	AND	*TotDist*	is	NOT High
Then	State	is	Undervoltage				

APPENDIX 4 DIAGNOSTIC RESULTS FROM THE HW55S TURNOUT SYSTEM

A4.1 Diagnostic results of the training dataset

The confusion matrix of all diagnostic results obtained from the N-to-R training HW55S operations is illustrated in Figure A4-1, demonstrating the overall accuracy of 81.60%. Also, the associated fault occurring likelihood trends is shown in Figure A4-2.

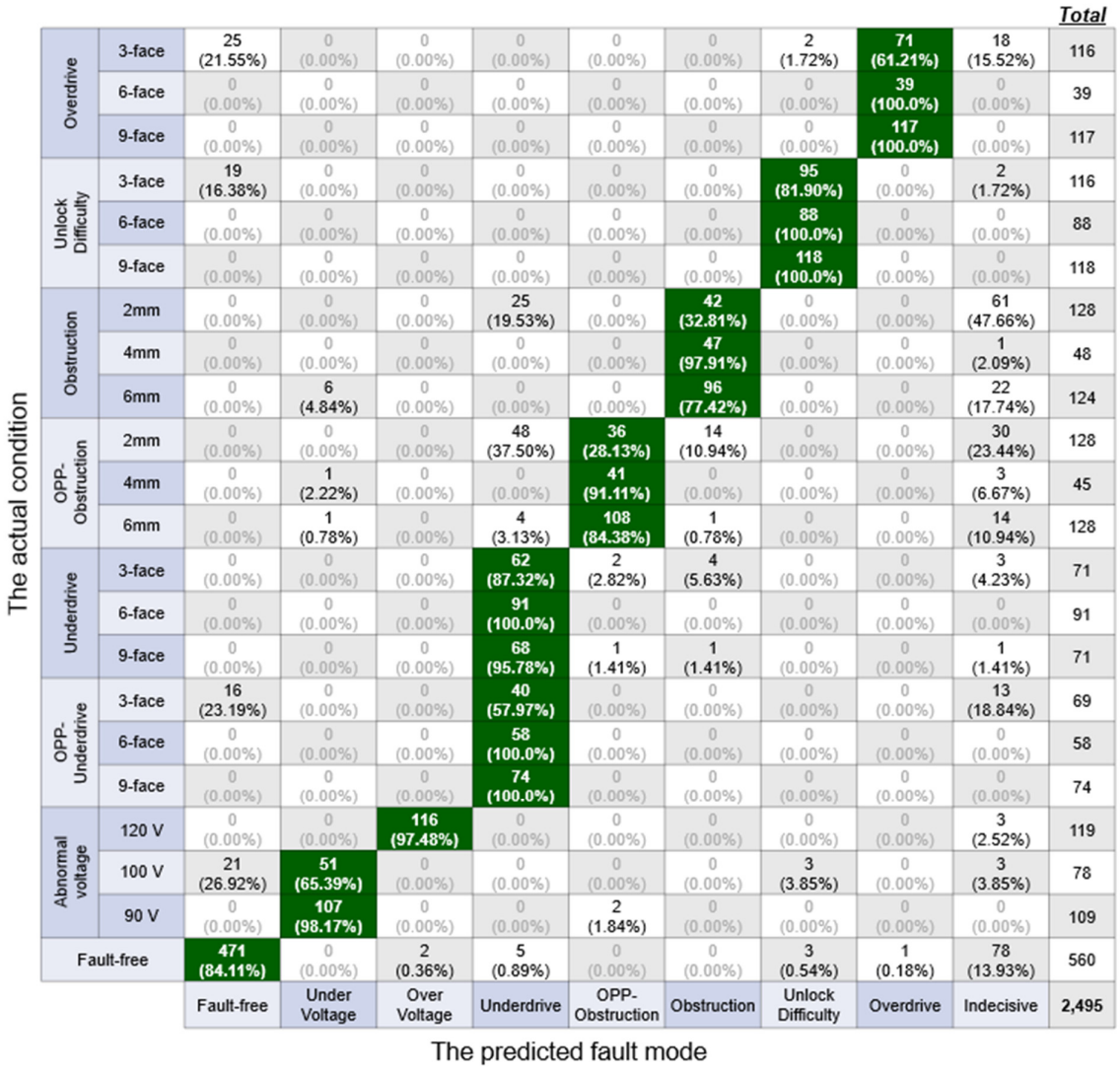


Figure A4-1 Confusion matrix showing the FDD results of the N-to-R HW55S turnout training dataset produced by the proposed methodology with the likelihood threshold at 0.5

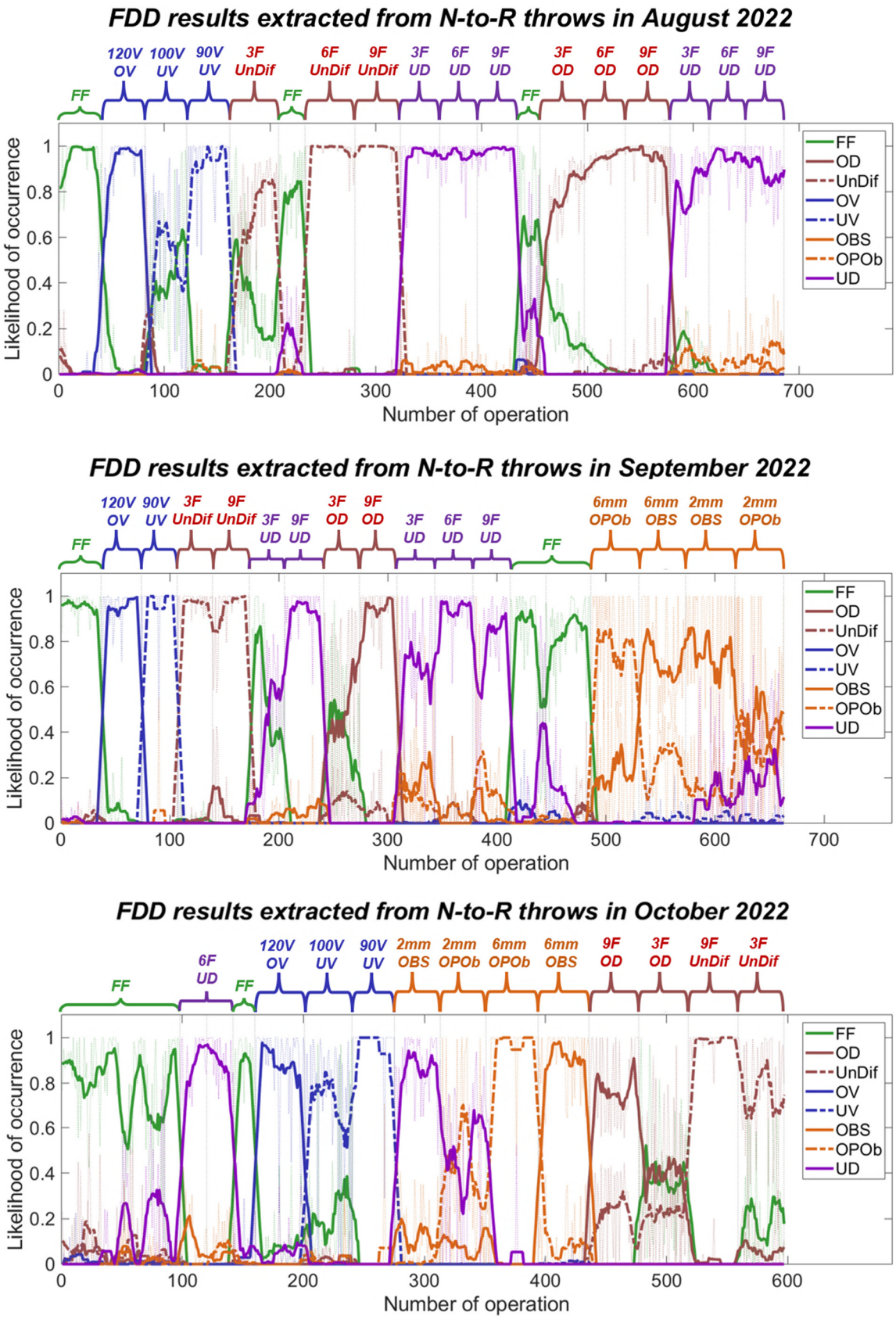


Figure A4-2 Examples of occurrence likelihood trends extracted from N-to-R training dataset

The confusion matrix of all diagnostic results obtained from the R-to-N training HW55S operations is illustrated in Figure A4-3, demonstrating the overall accuracy of 84.02%. Also, the associated fault occurring likelihood trends is shown in Figure A4-4.

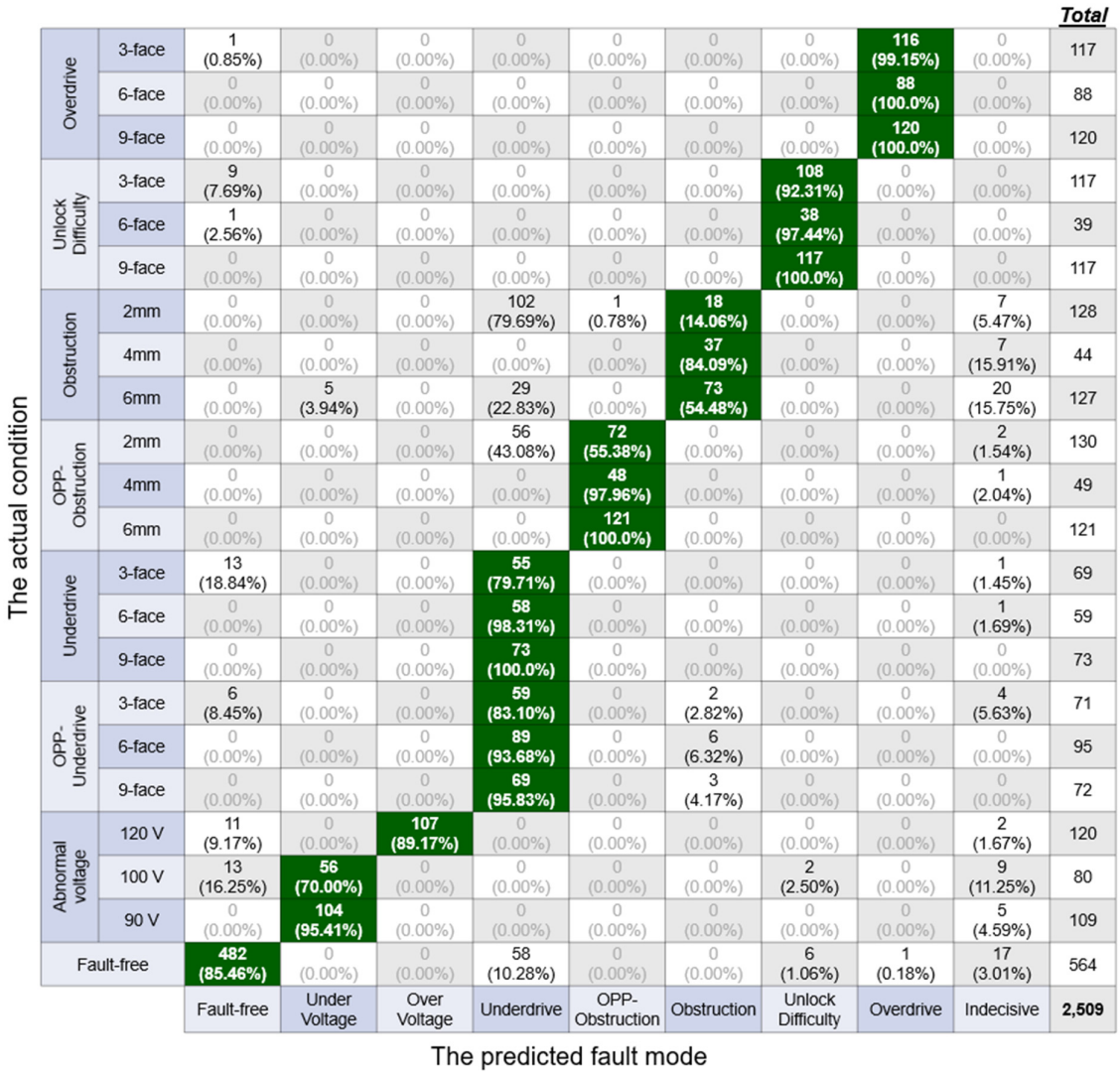


Figure A4-3 Confusion matrix showing the FDD results of the R-to-N HW55S turnout training dataset produced by the proposed methodology with the likelihood threshold at 0.5

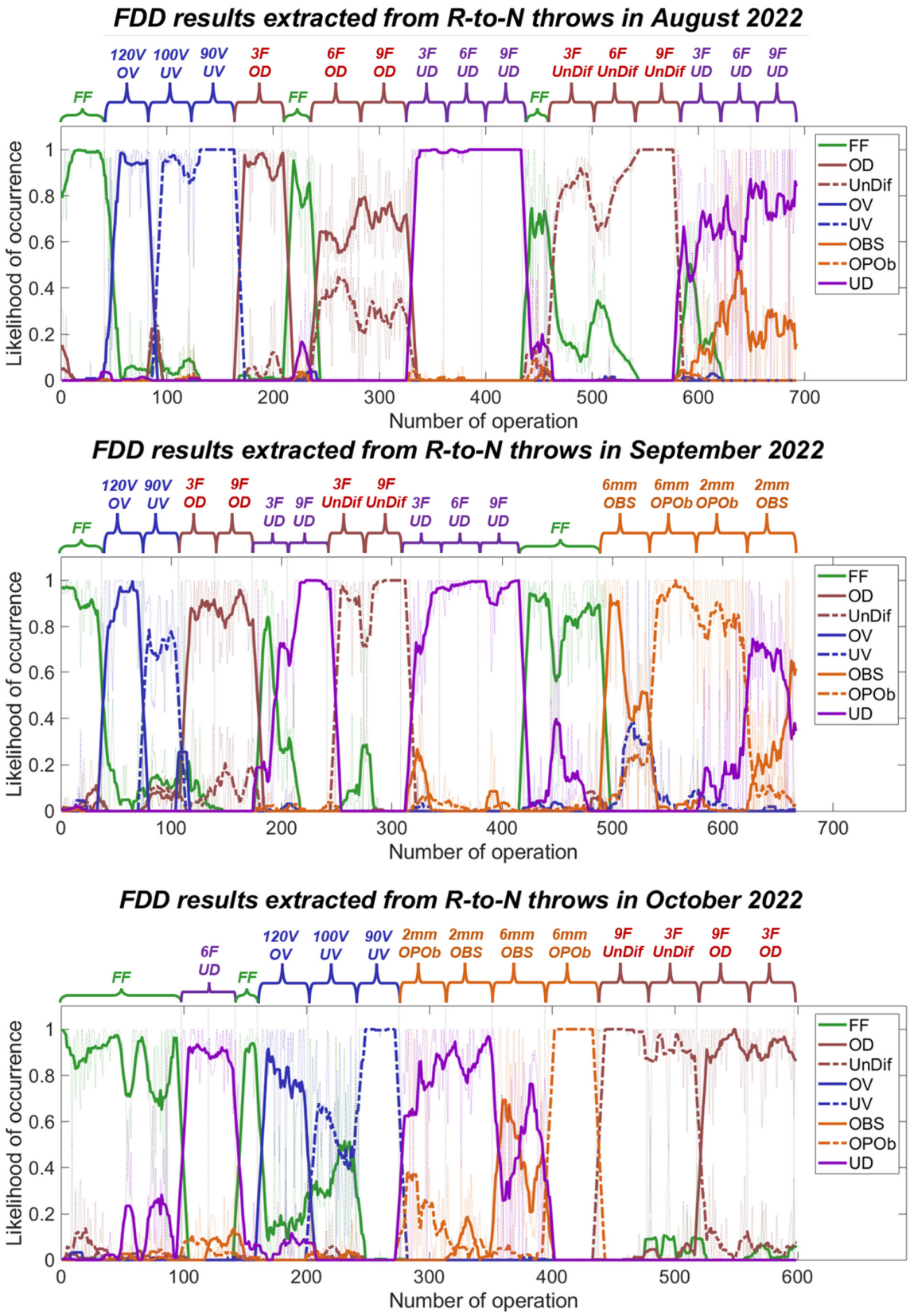


Figure A4-4 Examples of occurrence likelihood trends extracted from R-to-N training dataset

A4.2 Diagnostic results of the testing dataset

The confusion matrix of all diagnostic results obtained from the N-to-R testing HW55S operations is illustrated in Figure A4-5, demonstrating the overall accuracy of 84.05%. Also, the associated fault occurring likelihood trends is shown in Figure A4-6.

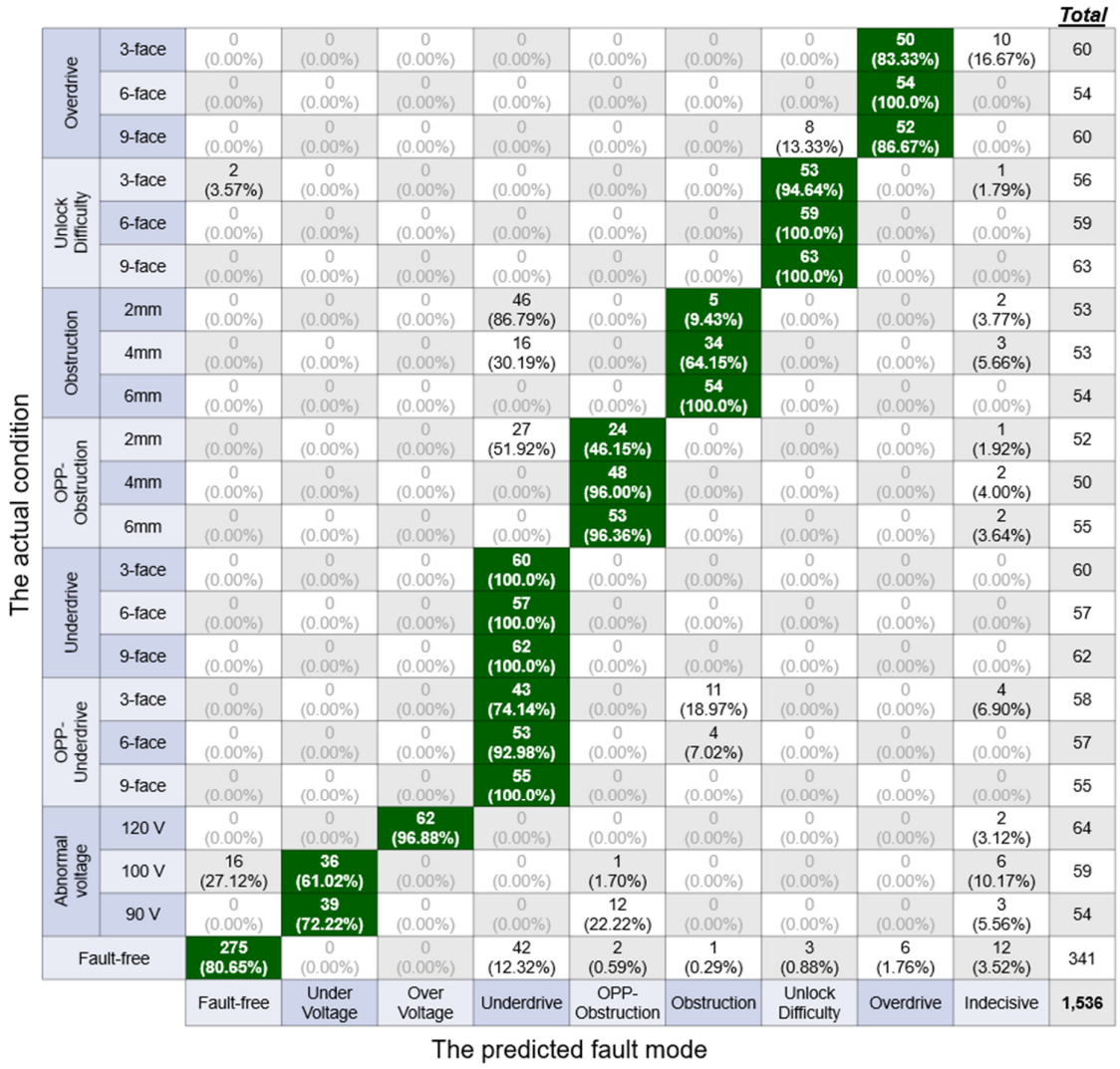


Figure A4-5 Confusion matrix showing the FDD results of the N-to-R HW55S turnout testing dataset produced by the proposed methodology with the likelihood threshold at 0.5

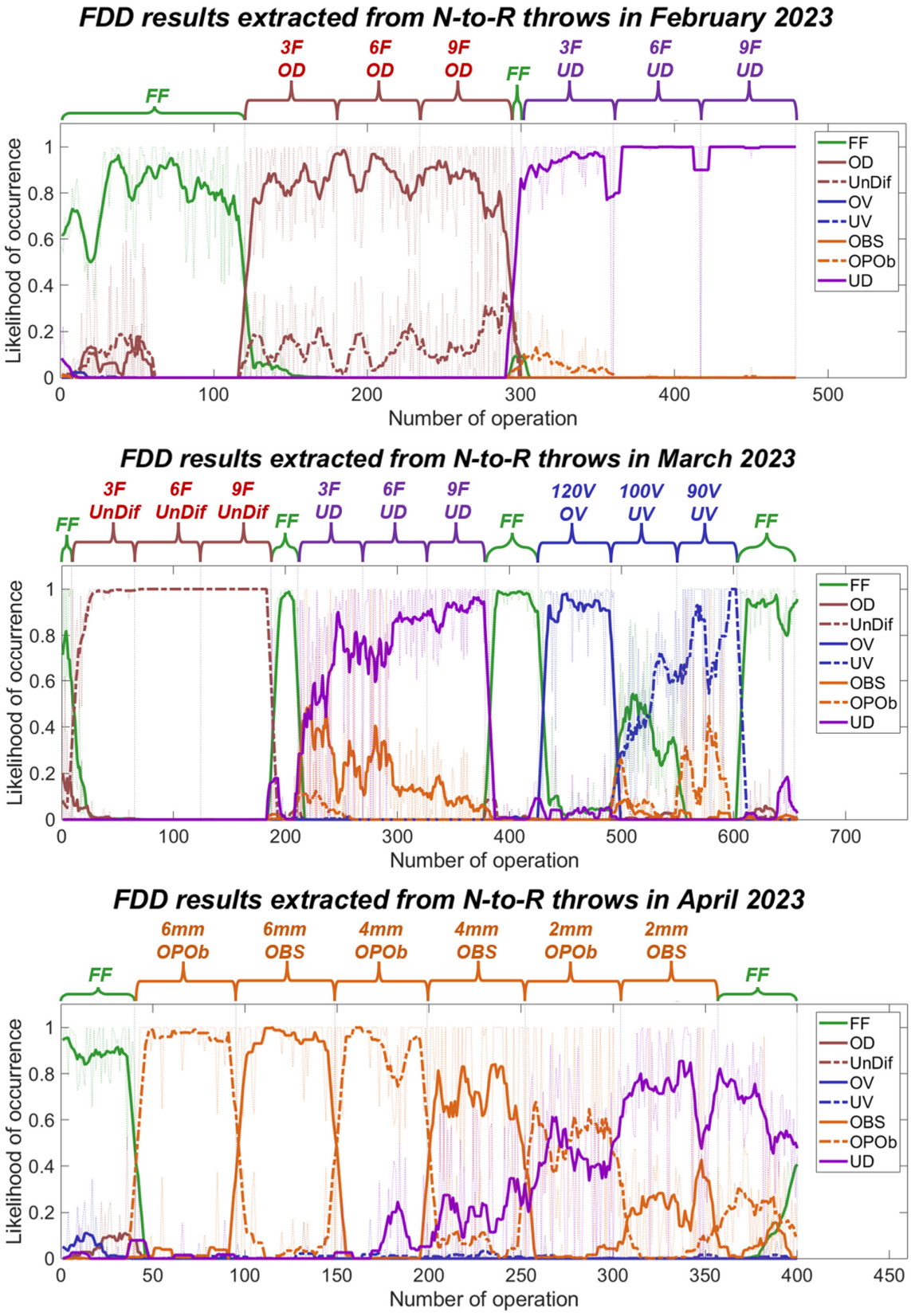


Figure A4-6 Fault occurrence likelihood trends of the entire N-to-R HW55S testing dataset

The confusion matrix of all diagnostic results obtained from the R-to-N testing HW55S operations is illustrated in Figure A4-7, demonstrating the overall accuracy of 83.71%. Also, the associated fault occurring likelihood trends is shown in Figure A4-8.

The actual condition		Fault-free	Under Voltage	Over Voltage	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	Total
Overdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	13 (22.81%)	**39 (68.42%)**	5 (8.77%)	57
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	37 (60.66%)	**24 (39.34%)**	0 (0.00%)	61
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (4.84%)	**59 (95.16%)**	0 (0.00%)	62
Unlock Difficulty	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**58 (100.0%)**	0 (0.00%)	0 (0.00%)	58
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**54 (100.0%)**	0 (0.00%)	0 (0.00%)	54
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**60 (100.0%)**	0 (0.00%)	0 (0.00%)	60
Obstruction	2mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	43 (82.69%)	0 (0.00%)	**8 (15.39%)**	0 (0.00%)	0 (0.00%)	1 (1.92%)	52
	4mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**47 (94.00%)**	0 (0.00%)	0 (0.00%)	3 (6.00%)	50
	6mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**53 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	53
OPP-Obstruction	2mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	49 (92.45%)	**1 (1.89%)**	1 (1.89%)	0 (0.00%)	0 (0.00%)	2 (3.77%)	53
	4mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (1.92%)	**51 (98.08%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	52
	6mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**56 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	56
Underdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**52 (88.14%)**	7 (11.86%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	59
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**57 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	57
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**56 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	56
OPP-Underdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**60 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	60
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**58 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	58
	9-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	**60 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	60
Abnormal voltage	120 V	0 (0.00%)	0 (0.00%)	**61 (95.31%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (4.69%)	64
	100 V	1 (1.75%)	**51 (89.47%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	5 (8.77%)	57
	90 V	0 (0.00%)	**19 (33.93%)**	0 (0.00%)	0 (0.00%)	26 (46.43%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	11 (19.64%)	56
Fault-free		**301 (88.53%)**	0 (0.00%)	0 (0.00%)	29 (8.53%)	0 (0.00%)	1 (0.29%)	1 (0.29%)	4 (1.18%)	4 (1.18%)	340
		Fault-free	Under Voltage	Over Voltage	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	**1,535**

The predicted fault mode

Figure A4-7 Confusion matrix showing the FDD results of the R-to-N HW55S turnout testing dataset produced by the proposed methodology with the likelihood threshold at 0.5

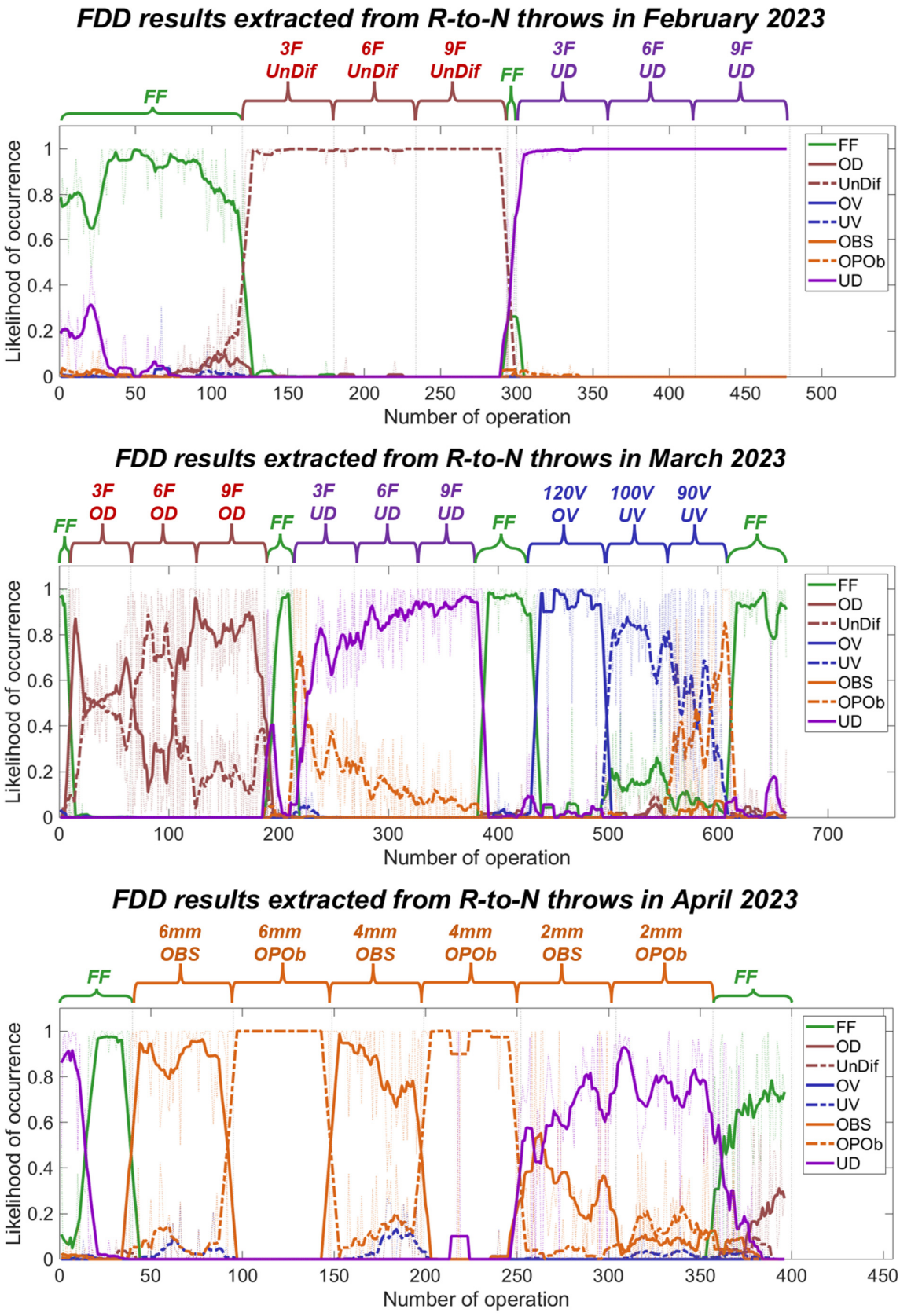


Figure A4-8 Fault occurrence likelihood trends of the entire R-to-N HW55S testing dataset

APPENDIX 5 DIAGNOSTIC RESULTS FROM THE HW2000 TURNOUT SYSTEM

A5.1 Diagnostic results of the N-to-R testing dataset

The confusion matrix of all diagnostic results obtained from the N-to-R testing HW2000 operations once the fuzzy rules have been modified is shown in Figure A5-1, demonstrating the overall accuracy of 87.55%. Meanwhile, the corresponding fault occurring likelihood trend is also plotted in Figure A5-2.

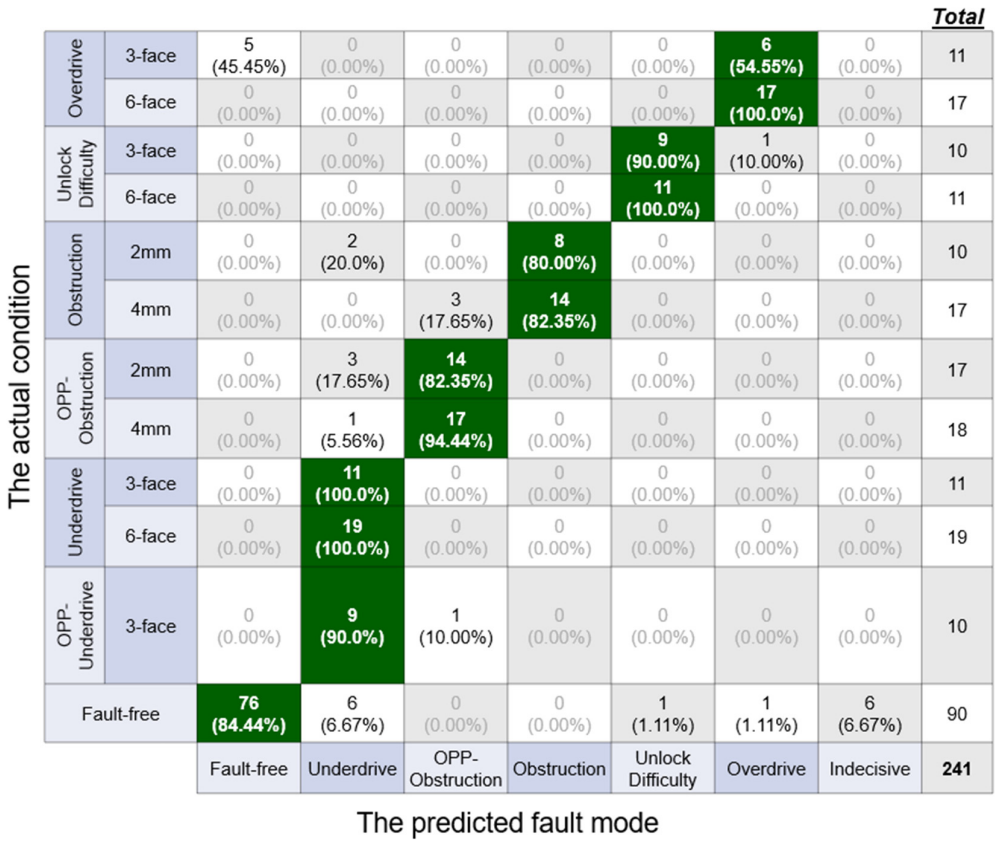


Figure A5-1 Confusion matrix showing the FDD results of the N-to-R HW2000 turnout testing dataset once the methodology has been modified to suit the HW2000 turnout

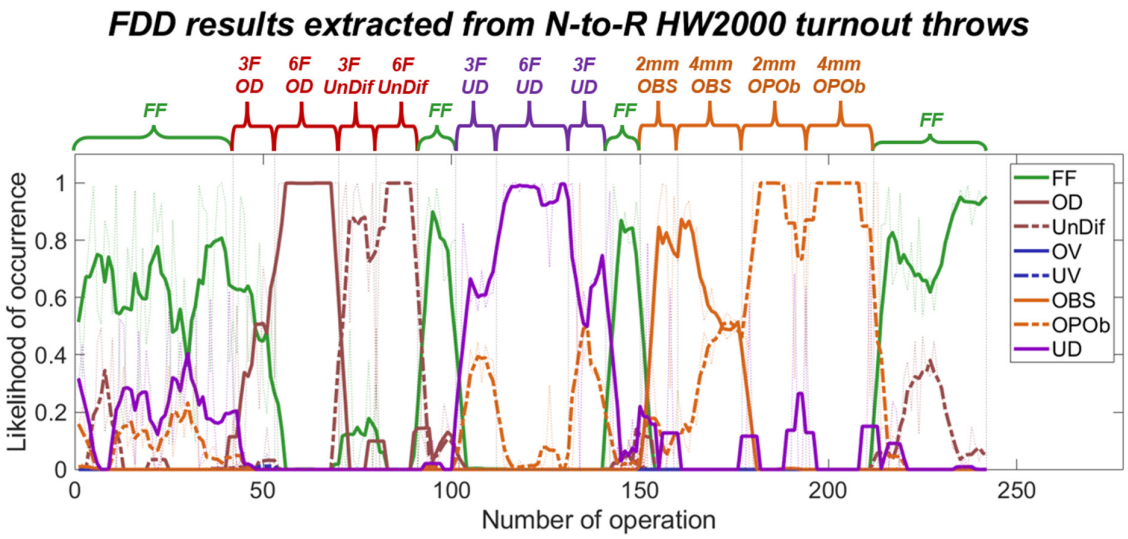


Figure A5-2 Fault occurrence likelihood trends of the entire N-to-R HW2000 testing dataset once the methodology has been modified to suit the HW2000 turnout

A5.2 Diagnostic results of the R-to-N testing dataset

The confusion matrix of all diagnostic results obtained from the R-to-N testing HW2000 operations once the fuzzy rules have been modified is shown in Figure A5-3, demonstrating the accuracy of 72.53%. Meanwhile, the corresponding fault occurring likelihood trend is also plotted in Figure A5-4. In this scenario, the inferior accuracy in diagnosing R-to-N operations (compared to 87.55% in N-to-R operations) is expected from the asymmetry of the HW2000 turnout system, causing the effects of the same fault simulations to be different. This results in less severe faulty operations being misclassified by the fuzzy rule-based FDD algorithm.

The actual condition		Fault-free	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	Total
Overdrive	3-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (10.00%)	**4 (40.00%)**	5 (50.00%)	10
	6-face	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**10 (100.0%)**	0 (0.00%)	10
Unlock Difficulty	3-face	7 (70.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**2 (20.00%)**	0 (0.00%)	1 (10.00%)	10
	6-face	1 (10.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**9 (90.0%)**	0 (0.00%)	0 (0.00%)	10
Obstruction	2mm	0 (0.00%)	10 (58.82%)	5 (29.41%)	**1 (5.88%)**	0 (0.00%)	0 (0.00%)	1 (5.88%)	17
	4mm	0 (0.00%)	0 (0.00%)	0 (0.00%)	**17 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	17
OPP-Obstruction	2mm	0 (0.00%)	10 (90.91%)	**1 (9.09%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	11
	4mm	0 (0.00%)	3 (17.65%)	**14 (82.35%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	17
Underdrive	3-face	0 (0.00%)	**11 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	11
OPP-Underdrive	3-face	0 (0.00%)	**10 (100.0%)**	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	10
	6-face	0 (0.00%)	**17 (94.44%)**	1 (5.56%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	18
Fault-free		**73 (79.35%)**	4 (4.35%)	1 (1.09%)	0 (0.00%)	6 (6.52%)	0 (0.00%)	8 (8.70%)	92
		Fault-free	Underdrive	OPP-Obstruction	Obstruction	Unlock Difficulty	Overdrive	Indecisive	**233**

The predicted fault mode

Figure A5-3 Confusion matrix showing the FDD results of the R-to-N HW2000 turnout testing dataset once the methodology has been modified to suit the HW2000 turnout

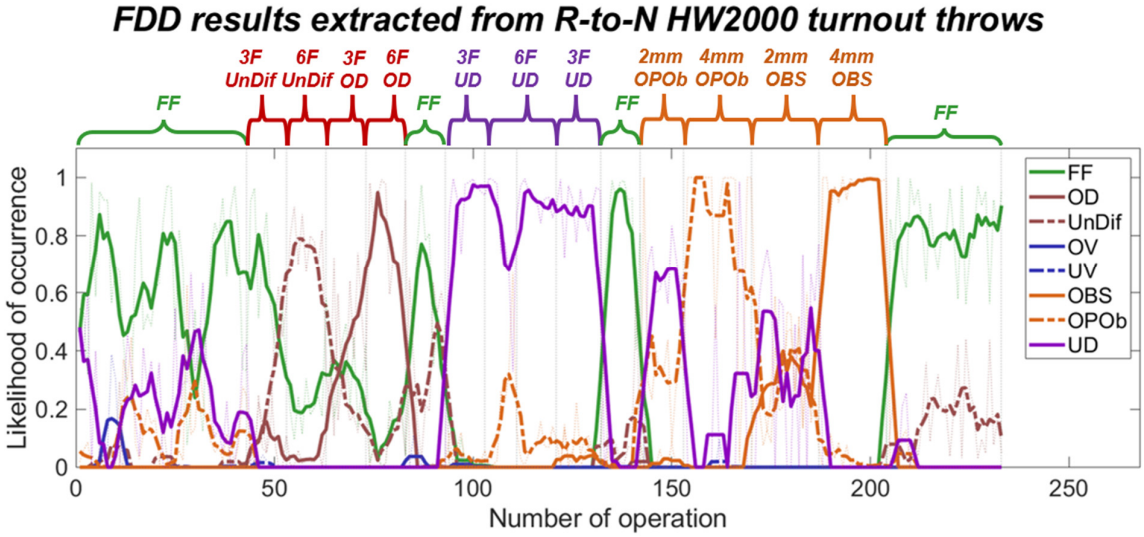


Figure A5-4 Fault occurrence likelihood trends of the entire R-to-N HW2000 testing dataset once the methodology has been modified to suit the HW2000 turnout